VISUAL GUIDANCE FOR A MOBILE ROBOT USING THE GENERALIZED SHIFTED RECIPROCAL-WEDGE TRANSFORM

by

Simon X. O. Ren

B.Sc. Simon Fraser University 1994

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
April 1996

APPROVAL

Name: Simon X. O. Ren

Degree: Master of Applied Science

Title of thesis: Visual Guidance for a Mobile Robot Using the Generalized Shifted Reciprocal-Wedge Transform

Examining Committee: Dr. Jamal Deen
Chair

Dr. William A. Gruver (Senior Supervisor)
Professor, Engineering Science

Dr. Ze-Nian Li (Senior Supervisor)
Associate Professor, Computing Science

Dr. Kamal Gupta (Supervisor)
Associate Professor, Engineering Science

Dr. John Jones (Thesis Examiner)
Associate Professor, Engineering Science

Date Approved: APRIL 15, 1996

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 usrs. 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

"Visual Guidance For a Mobile Robot Using The Generalized Shifted Reciprocal Wedge Transformation"

Author:

_____
(signature)

X. Ren
(name)

April 15th, 1996
(date)

Abstract

Mobile robot navigation relies heavily on landmark recognition. In a corridor environment notable objects and signs are often on the surrounding surfaces such as walls and doors. Since ordinary camera views may be perspectively distorted, and the degree of distortion varies greatly due to movements of the camera, object recognition by matching a simple, uniform object model is not feasible.

This thesis develops an image transformation — the Generalized Shifted Reciprocal-Wedge Transform (*GS-RWT*) for compensating perspective distortions. The *GS-RWT* is based on the Reciprocal-Wedge Transform (*RWT*) space-variant image representation and its variant, the Shifted Reciprocal-Wedge Transform (*S-RWT*). It is characterized by three parameters δ, a and γ. First, it is shown that the distortion caused by camera tilt can be compensated by the use of the *S-RWT*. The shift constant a of the *S-RWT* is determined by the camera tilt angle. Second, the Generalized *S-RWT* (*GS-RWT*) is developed to compensate the distortion caused by both the camera pan and tilt. It is shown that the parameters (δ, a and γ) of the *GS-RWT* can be derived from the camera pan and tilt angles. The resulting image is a canonical frontal view of one of the surrounding surfaces. Therefore, a unique and uniform object model can be used for object recognition, which significantly simplifies

recognition task.

The *GS-RWT* has the property that the space-variant resolution occurs primarily in one direction which enables the preservation of the original shape of the object. Experimental results from several robot corridor navigation examples are presented to demonstrate the application of the *GS-RWT* to camera gaze control and object tracking and approaching. Gaze control of the camera is simplified because the pan-tilt angles can be directly measured from the canonical frontal view. Object tracking and approaching is facilitated since object recognition is performed on the recovered uniform resolution images.

Contents

List of Figures

Chapter 1

Introduction

Mobile robot navigation is a remarkably challenging task. Beside the mechanical and electrical challenges, the robot must always answer the following questions: "where am I?", "where am I going?", and "how should I get there?" [LDW92]. Under some simplified conditions, the robot could rely on prescribed maps to follow some pre-planned paths, or use various non-vision sensors to fulfill the tasks. In general, robot navigation without vision has been proven very difficult and limited. Often, the best demonstrated capability of the robot is to wander around by hastily avoiding obstacles.

In seeking assistance from computer vision, many researchers have considered the model-based vision — an approach to visual object recognition [Low87, CMST87, Gri90]. The aim of the model-based vision is to recognize an object in an environment and determine the position and orientation of this object relative to the robot. Model-based approaches usually require that geometrically precise models of a 3D environment be available, which distinguishes these approaches from the manner in which humans navigate. Meng and Kak [MK93] point out the following characteristics

of human navigational behavior. "First, human navigators do not need to calculate the exact coordinates of their position while navigating roads or hallways. Secondly, for a given goal, human navigators focus attention on particular stimuli from their visual input and very quickly extract meaningful information. Finally, extra information may be extracted from the scene during reactive behavior; this information will usually be stored and may be retrieved subsequently for higher level reasoning." In their work, instead of the traditional geometrical representation of the environment, a semantically richer, symbolic representation is used in which a hall way is modeled by the order of appearance of various landmarks and by adjacency relationships. Robot navigation is guided by the visual landmark detection.

Aloimonos [AWB88, Alo90] argued that we can achieve many highly non-trivial visual tasks in navigation without solving the general structure from motion problem. Robot systems should purposively look for important features which are essential to the navigation tasks. The statement, "I move, therefore I see" [Ham92], describes this situation. An active observer is one capable of engaging in an activity and able to control the geometric parameters of its sensory apparatus [Alo93]. Compared with a passive observer, who is not capable of choosing how to view a scene but is instead limited to what is offered, an active observer works in a much more efficient and intelligent way.

1.1 Active Vision

The area of active vision has attracted much attention in recent years [AWB88, Baj88, Bal91, Tso92, SS93]. It is argued that perception is not a passive process, but rather an active process of exploratory, probing and search. An active system differs from a

passive system in its purposive interaction with the world.

Aloimonos [Alo90] studies purposive vision. It is argued that computer vision should not be studied by itself, but in the context of a big process in which vision is used as help. Such a vision system is defined according to the task as a collection of processes each of which is to solve a particular sub-task related to the original visual problem. Very often, these sub-tasks are simple enough that they require only a qualitative decision from the visual process.

The term "active sensing" is used by Bajcsy [Baj88] to define a data acquisition process which adapts to the current data and the goal of the task. A visual system in this perspective encompasses both local and global models of the 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.

Krotkov [Kro93, Kro89] demonstrated an active vision system which combines several range recovery techniques, so that they cooperate in visual behaviors similar to those exhibited by the human visual system. Implemented features include aperture adjustment, focus ranging followed by fixation, stereo ranging followed by focus ranging, etc.

Ballard [Bal91] argued that vision is in the context of visual behaviors that the system is engaged in. One important feature of active vision is gaze control. Gaze control is to actively manipulate the imaging system in order to acquire images that are directly suited to the tasks to be performed [SS93]. Without it the visual system must work in isolation, with the burden of solving difficult problems with many degrees

Table 1.1: Categorization of model match

Image Portions	Object to Match Against	
	One	Many
One		Identification
Many	Location	Image interpretation

of freedom. Gaze control is gained by the interrelation between the visual processing and sensory-motor behaviors. A moving camera under ego-motion provides additional constraints on the image process [AWB88].

A real-time vision requires a system that purposively gathers information from the visual world. Useful information in the world is often distributed in space. Landmark recognition is a model matching problem. A landmark model is usually complicated. However, it can often be spatially divided into several simple sub-models. The general image interpretation by associating many models to many parts of the image in order to achieve the landmark recognition is too hard [SS93]. Due to the cost and complexity considerations, the camera is required to focused on restricted regions of the scene. In order to make it computationally efficient within a single camera position, interpretation problem has to be simplified, either into one of location or identification problem, as summarized in Tab. 1.1 [SS93]. In our landmark recognition task, gaze control is utilized to focus the camera on one single sub-object at a time and object tracking and approaching is performed to fixate the object and gain a better view during the motion. The system tries to match the sub-object with a sub-model which is usually easy to achieve. Therefore, it turns a "Many—Many" image interpretation problem into a "One—Many" object location problem.

1.2 Space-variant Sensing

It has been argued [Sch77, SD90, SS93] that spatially variable-resolution sensors are central to the sensing mechanism of an active vision system because they are economic and effective when coupled with active control. 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 human visual system also 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.

Ballard [Bal89, Bal91] uses the term *animate systems* as a contrast to passive systems. In his work, it has been shown that visual computation can be vastly less expensive when considered in the larger context of behavior. The most important visual behavior is the ability to control the direction of gaze which allows the use of very low resolution imaging that has a high virtual resolution. In addition, animate systems under real-constraints can further reduce their computational burden by using environmental cues that are perspicuous in the local context.

Space-variant imaging has been used for robot navigation. Bishay, Peters and Kawamura [BPK94] described an approach for indoor object detection using the Log-Polar transform. The image center for the log-polar mapping was selected to be the vanishing point of the corridor scene. The varying resolution property of the mapping simplified the model-matching so that a unique template can be used for every object. They also claimed that the detection was very fast due to the small size of the log-polar

images.

Tong and Li [TL93, Ton95] present an alternative image representation, the *Reciprocal-Wedge Transform* (RWT), to the log-polar transform for the space-variant imaging. It is shown that the RWT is applicable to road navigation, motion stereo, and ego-motion. The RWT is capable of compensating the perspective distortion while reducing the unnecessary details from its variable-resolution imagery [Ton95, LTR95], which renders it suitable for various robotics applications.

1.3 Objectives of this thesis

This thesis will examine certain aspects of the problem of visual guidance for robot navigation in simple indoor environments such as a corridor or an office. In an ordinary corridor scene, notable objects and signs are often on the surrounding surfaces, e.g., walls and doors. Since ordinary camera views may be perspective distorted, and the degree of distortion varies greatly due to movements of the camera, object recognition by matching a simple, uniform object model is not feasible.

This thesis develops an image transformation — the Generalized Shifted Reciprocal-Wedge Transform (GS-RWT) for compensating perspective distortions. The GS-RWT is based on the Reciprocal Wedge Transform (RWT) space-variant image representation and its variant, the Shifted Reciprocal Wedge Transform (S-RWT). It is characterized by three parameters δ, a and γ. First, it is shown that the distortion caused by camera tilt can be compensated by use of the S-RWT. The shift constant a of the S-RWT is determined by the camera tilt angle. Second, the Generalized S-RWT (GS-RWT) is developed to compensate the distortion caused by both the camera pan and tilt. It is shown that the parameters (δ, a and γ) of the GS-RWT

can be derived from the camera pan and tilt angles. The resulting image is a canonical frontal view of one of the surrounding surfaces. Therefore, a unique and uniform object model can be used for object recognition, which significantly simplifies the recognition task.

The *GS-RWT* has the property that the space-variant resolution occurs primarily in one direction which enables the preservation of the original shape of the object. Experimental results from several robot corridor navigation examples are presented to demonstrate the application of the *GS-RWT* to camera gaze control and object tracking and approaching. Gaze control of the camera is simplified because the pan-tilt angles can be directly measured from the canonical frontal view. Object tracking and approaching is facilitated since object recognition is performed on the recovered uniform resolution images.

The organization of the rest of the thesis is as follows: Chapter 2 gives a brief review of the *RWT* model and its properties. Chapter 3 describes a method based on *S-RWT* to compensate the perspective distortion caused by camera tilt. This approach is extended in Chapter 4 and the *GS-RWT* is introduced to compensate the perspective distortion caused by both camera pan and tilt. The parameters of the *GS-RWT* is shown to be derived from camera pan-tilt angles. Chapter 5 describes the application of the *GS-RWT* in camera gaze control and object tracking and approaching in robot navigation. Chapter 6 presents the conclusions and discusses the potential extensions for future researches.

Chapter 2

RWT Review

2.1 Reciprocal-Wedge Transform

A brief review of the Reciprocal-Wedge Transform (RWT) is given in this chapter[1]. The RWT was proposed as an alternative model for space-variant sensing [TL93, TL95b]. Different from log-polar transform [WC79, SD90] where space-variant resolution is achieved by centric scaling, the resolution variation of the RWT is anisotropic, predominantly in one dimension. Consequently, the RWT preserves linear features in the original image. This makes the transform especially suitable for vision applications that are related to linear structures or are translational in nature, such as line detection, linear motion and stereo correspondence. The RWT has been applied in road navigation and motion stereo [TL94, LTR95].

[1]The figures in this chapter are reprinted with the permission of the authors.

2.1.1 Introduction to the RWT

The *RWT* maps a rectangular image into a wedge shaped image (see Fig. 2.1). Space-variant resolution is achieved as the smaller side of the wedge is sampled with fewer pixels than the wider side is. Mathematically, the *RWT* is defined as a mapping of the image pixels from the x-y domain to a u-v domain such that

$$u = 1/x \ , \quad v = y/x \ . \tag{2.1}$$

The result of the *RWT* is the introduction of a reciprocal transformation. The transformed image shows a wedge-shape in an inside-out fashion because of the scaling effect of the x reciprocal function. As illustrated in Fig. 2.1, 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. This results in an anisotropic space-variant resolution.

Due to the property of anisotropic space-variant resolution, the *RWT* preserves the linearity of lines in the x-y domain. Let L_{xy} and L_{uv} be the representations of a line in x-y and u-v domain. By Eq. (2.1), the following transformation dual of a line can be derived:

$$L_{xy} : y = m \cdot x + c \ , \quad L_{uv} : v = c \cdot u + m \ . \tag{2.2}$$

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 linearity is preserved. It is interesting to note that the values for the slope and intercept are interchanged between the transformation. We infer that parallel lines with the same slope in x-y will be mapped to u-v lines having

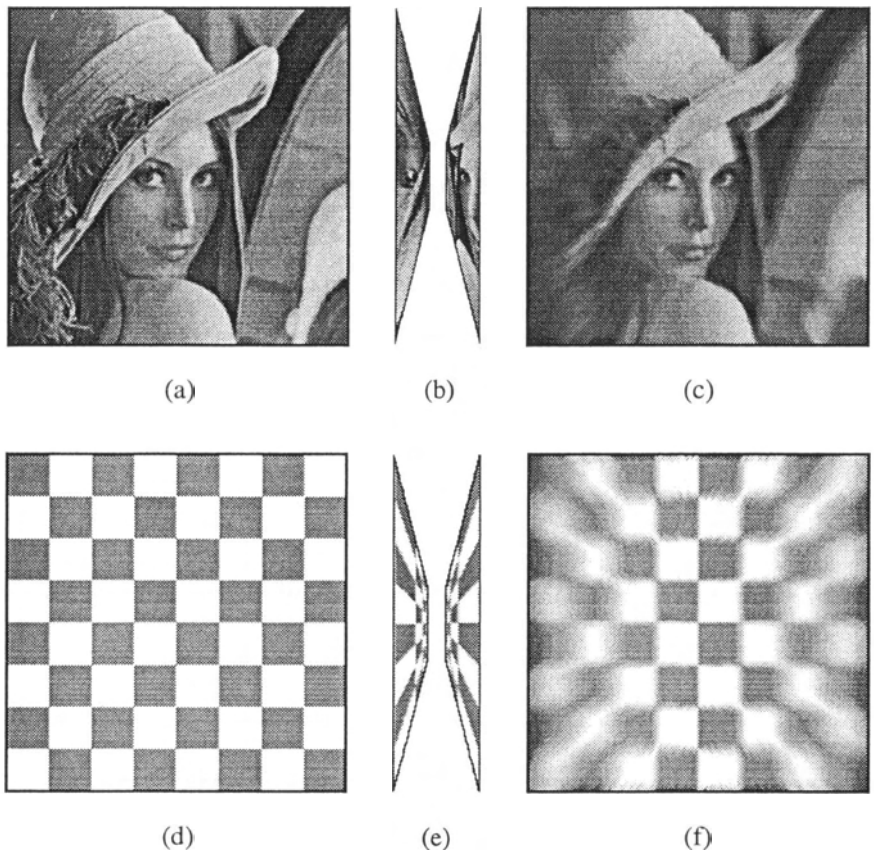


Figure 2.1: The Reciprocal-Wedge transform.
(a) The lady's image. (b) The *RWT* image shows two inside-out wedges. (c) The image when transformed back to the Cartesian domain. (d) A rectangular grid. (e) The *RWT* image. (f) The grid transformed back to illustrate the resolution varying from the center to the periphery.

the same v-intercept. Inversely, lines having the same y-intercept will form parallel lines in the u-v domain.

2.1.2 Matrix notation of the RWT

The RWT defined in Eq. (2.1) can be represented as a matrix notation. By using the homogeneous coordinates, the RWT can represented as a cross-diagonal matrix of 1's, as shown in Eq. 2.3; and the inverse of the RWT matrix $\mathbf{T}$ is $\mathbf{T}$ itself.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{T}^{-1} , \tag{2.3}$$

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

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

This concise matrix notation simplifies the use of the RWT with other geometrical transformations. Various transformations can be performed easily in Cartesian or RWT domains. The discussion of the Generalized Shifted Reciprocal-Wedge Transform in later chapters takes advantage of this property.

2.1.3 Projective Model of the RWT

The RWT can be represented as a projection of an image on a plane perpendicular to it. Consider the perspective projection in which the three-dimensional XYZ space is projected onto the two-dimensional Z-Y plane at $X = 1$ (Fig. 2.2). Let the three-dimensional point be (X, Y, Z) and the projection be (Z', Y'), then

$$Z' = Z/X , \quad Y' = Y/X . \tag{2.4}$$

Eq. (2.1) is equivalent to Eq. (2.4), if the terms $x, y, 1, u, v$ in Eq. (2.1) are unified with the X, Y, Z, Z', Y' in Eq. (2.4), correspondingly. Therefore, the *RWT* described by Eq. (2.1) can 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 $X = 1$.

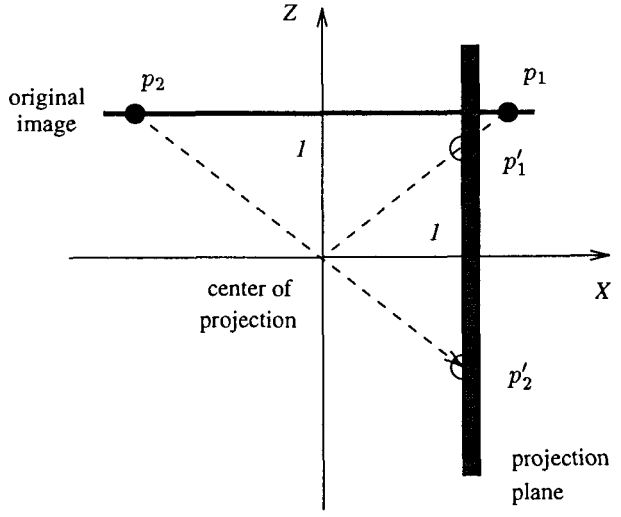


Figure 2.2: A perspective projection model.
The original image is placed on the X-Y plane at $Z = 1$. It is reprojected onto the Z-Y projection plane at $X = 1$. The pixels p_1 and p_2 are projected to p'_1 and p'_2, respectively.

2.2 Shifted Reciprocal-Wedge Transform

2.2.1 Singularity avoidance

The singularity of the *RWT* exists at $x = 0$, i.e., $u = 1/0 = \infty$ and $v = y/0$ (see Eq. (2.1)). The Shifted Reciprocal-Wedge Transform (*S-RWT*) [TL93] is first introduced to avoid singularity[2]. The *S-RWT* is a variation of the *RWT* with a shift parameter

[2]Another method, so called, patching method [Ton95], is also used to solve this problem.

a in the x dimension.

$$u = 1/(x+a)\,, \quad v = y/(x+a)\,. \tag{2.5}$$

Both the forward and backward transformations for the *S-RWT* have the same cross-diagonal matrix Eq. (2.3) except for the presence of the additional parameter a.

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

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 shifted in opposite directions.

2.2.2 S-RWT and V-plane projection

The *S-RWT* discussed in Section 2.2.1 can be implemented with a V-plane projection. Fig. 2.3 depicts the V-plane projection. The two projection planes are joined to form a V shape 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. It can be observed that the orientation of the V arms is not as steep as that of the 90° projection plane in Figure 2.2 which is the projective model for *RWT*. A less drastic space-variant resolution should be expected.

Tong [Ton95] showed that the V-plane projection implements the space-variant resolution of the *S-RWT* of Eq. (2.5). The shift parameter a depends upon the

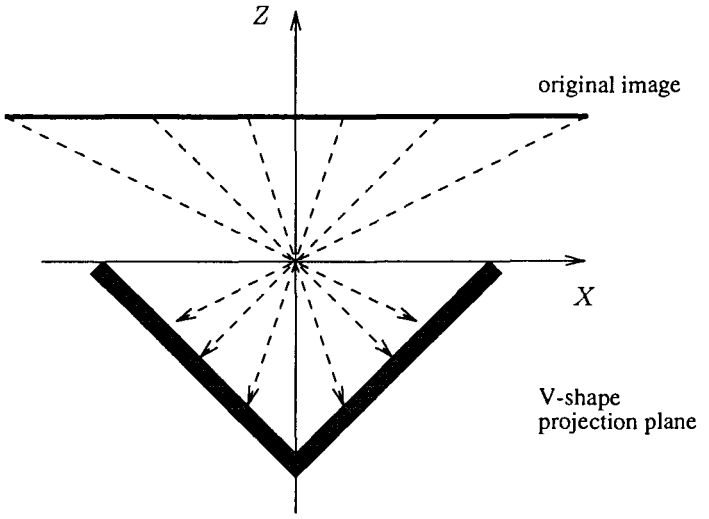


Figure 2.3: V-plane projection.
The left arm of the V forms the projection plane for the right half of the original image and the right arm of the V is for the left half. The singularity problem is resolved, and space-variant resolution is effected on both projection planes.

V-plane interception angle. As illustrated in Fig. 2.4, the mapping of a point P on the original image plane to the projection point Q on the projection is following:

$$u = \frac{f^2}{x+a}, \tag{2.6}$$

$$v = \frac{fy}{x+a}. \tag{2.7}$$

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

From Eq. (2.6) and Eq. (2.7), it is shown that the u and v coordinates from the V-plane projection are effectively computing the $S\text{-}RWT$ as defined in Eq. (2.5) with a constant factor f. This result bridges the relation between V-plane projection angle and the shift constant a, and it will be generated in the discussion in later chapters.

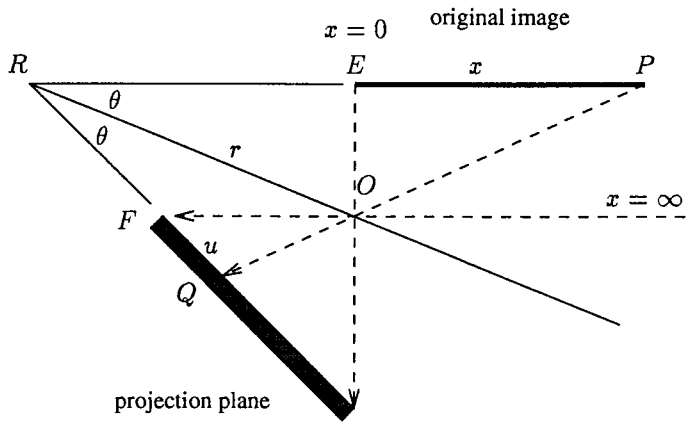


Figure 2.4: Geometry of the V-projection from P to Q.

Chapter 3

Distortion Compensation by the S-RWT

3.1 Robot Navigation and Perspective Distorted Views

Mobile robot navigation requires interaction with the surrounding environment in order to position the robot and decide its subsequent behaviors. We auned to design a robot to deliver small objects (mail, food, medicine) to rooms in office buildings and hospitals. The working space is a closed area formed by planar structures, such as walls, floor and ceiling.

The sensory system of the mobile robot consists of a CCD video camera and a ring of 16 ultrasonic sensors. Visual feedback plays the major role in this application. Natural landmarks are selected to guide the navigation or trigger certain behaviors. For example, an office room is found by detecting the door; characteristic features,

such as the door knob and the lock, are recognized and serve as intermediate objects for the indirect search [WB94] of the door. Most of the landmarks share the common features that are natural to a human navigator and geometrically simple to represent. In this thesis, landmarks are considered to be $2D$ coplanar patterns. Models for predefined landmark features are available. The recognition is hence restricted to be a $2D$ template matching problem. The $2D$ plane which contains the landmarks is called the *object plane*. Object planes are physical walls, floor or ceiling. Walls can be represented as continuous segments of vertical planes. Ultrasonic sensors are used to estimate the distance to walls and the orientation of the wall.

The capability of actively and purposively searching for landmarks enables the robot to focus attention on particular stimuli from its visual input and very quickly extracting meaningful information [MK86, AWB88, Alo90]. The mobile robot system is capable of translational and rotational movements and rotational movements for the pan-tilt mechanism on which camera is mounted. By controlling the geometric parameters of its sensory apparatus, the robot is capable of choosing how to view a scene, and, therefore, works in a much more efficient and intelligent way. On the other hand, images taken from arbitrary view points in the $3D$ world complicate the $2D$ landmark matching.

Assuming a standard pin-hole camera model (see Section 3.2), the image projection of the world point (X, Y, Z) is the image point $(X/Z, Y/Z)$. In this case, the appearance of any $3D$ object is governed only by the relative position and orientation of the camera with respect to the object, i.e., the camera pose. There are 6 degrees of freedom for the camera pose: three for rotation and three for translation. Projections of $2D$ coplanar landmarks vary with different configurations of these 6

parameters. The distortion of the image introduced by different camera pose is called the *perspective distortion*, due to the perspective projection model being used.

More restrictive transformations deserve special attention. Often these transformations are more readily computed, thus making matching easier. One such special case occurs for *frontal planes*, the object plane whose normal is parallel to the camera viewing direction. The perspective projection of a frontal plane to its appearance in the image can be described with four transformation parameters: an image rotation angle about camera viewing direction, a $2D$ translation vector, and an image scale [Saw92]. Therefore, the projection of the $2D$ object on the frontal plane is shape-preserving. The camera view with such characteristics is called a *frontal view.*

For frontal view, matching can be accomplished by standard model matching algorithm, such as the General Hough Transform [Bal81]. The General Hough transform is a voting algorithm in which the possible location of the object to be matched is supposed to get the most votes in the search space. Since there are four parameters remained for the frontal view, the Hough search space has four degrees accordingly: two for the $2D$ location on the image and one each for scale and rotation of the $2D$ template. Scale of the image is inversely proportional to the distance from camera to object plane. Search space can be further reduced by eliminating the rotation parameter. Such a view is called *canonical frontal view*, and the object plane which cause the canonical frontal view is called *canonical frontal plane.* Objects in the canonical frontal view should have the same orientation as their models which is usually represented canonically.

Due to the arbitrary camera pose, object plane is not always canonical frontal with respect to camera. The image of such a plane is perspective distorted. Fast

and reliable matching techniques exist when good initial guesses of camera pose or motion are available [BWR90, BR92]. What is lacking are good methods for finding matches in monocular images, formed by perspective projection, and taken from arbitrary viewpoints. In this chapter, a novel method based on the *Generalized Shifted Reciprocal Wedge Transform* is presented to convert the distorted image into a canonical frontal view of the object plane. Camera pose with respect to the object plane is assumed to be known, and is used to determine the transformation parameters. The resulting image therefore has three degrees of freedom: two for translation and one for scale. Scaling factor can be estimated by camera pose and the distance of the camera to object plane. The approximate distance can be read from the ultrasonic sensor. Therefore, the dimension Hough search space is reduced to three and is well bounded: $2D$ location is bounded by the image boundary and scaling factor is within a certain range of the estimated value.

3.2 Camera Model and Vanishing Point

3.2.1 Camera model

A pin-hole camera model is used throughout this thesis. As illustrated in Fig. 3.1, the coordinate system is right-handed and camera centered. The origin is set at the camera optical center, and the Z-axis is the camera viewing direction. The image plane is perpendicular to the Z-axis and intersects it at point $(0, 0, f)$, where f is the camera focal length. The x,y axes of the image coordinate are collimated with the camera X,Y axes. Using this model, a $3D$ point (X, Y, Z) is projected onto point (x, y) on the image whose image coordinate x, y are given by

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \tag{3.1}$$

In general, perspective projections of parallel 3D lines not perpendicular to the optical Z-axis meet at a *vanishing point* on the image projection plane. To see this, consider a line L represented by

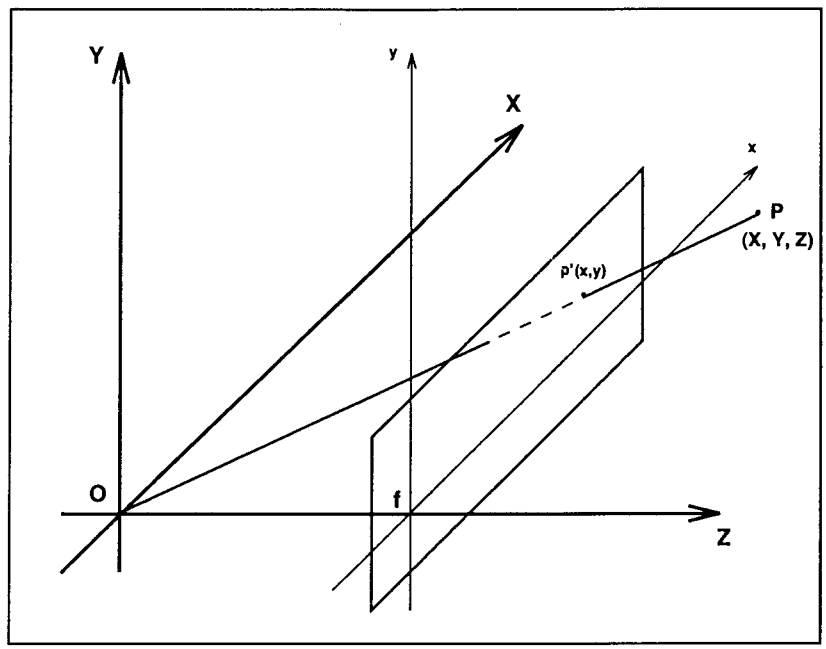


Figure 3.1: Pin-hole camera model

$$L = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \lambda \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{3.2}$$

where L passes through the point $(a_1\ a_2\ a_3)^T$ and has direction cosines $(b_1\ b_2\ b_3)^T$ with $b_3 \neq 0$, and λ a real number. From Eq. (3.1) and Eq. (3.2), the perspective projection $(x, y)^T$ of any point on the line L is given by:

$$x = f\frac{a_1 + \lambda b_1}{a_3 + \lambda b_3} \qquad y = f\frac{a_2 + \lambda b_2}{a_3 + \lambda b_3} \tag{3.3}$$

For lines with $b_3 = 0$, they are parallel to the image plane. The projection of these lines can be scaled, see Eq. (3.3). The result image is a frontal view. For parallel lines with $b_3 \neq 0$, their perspective projection lines converge to *vanishing points* in the image coordinate. These vanishing points are considered to be on the line and infinitely far from the center of the lens. Therefore, vanishing points can be computed as [HS93]:

$$x_\infty = \lim_{\lambda\to\infty} f\frac{a_1 + \lambda b_1}{a_3 + \lambda b_3} = f\frac{b_1}{b_3} \tag{3.4}$$

$$y_\infty = \lim_{\lambda\to\infty} f\frac{a_2 + \lambda b_2}{a_3 + \lambda b_3} = f\frac{b_2}{b_3}. \tag{3.5}$$

The convergence of parallel lines results in both reduction of size and nonuniform foreshortening of objects. Generally, the shape of an object is rarely preserved under a perspective projection. This is the cause of perspective distortion which has been mentioned in the previous section.

3.2.2 Pan-tilt model

The camera pan-tilt model is illustrated in Fig. 3.2. The camera is assumed to rotate about the origin O of the camera-centered coordinate system which is the camera optical center. For the initial setting, camera X-axis is vertical and camera Y-axis is horizontal[1]. Based on the characteristics of the pan-tilt device, camera

[1]Such a setting is used as a reference coordinate for the derivation of the *S-RWT* and the *GS-RWT*. X-axis has been used as the principle direction of the space-variation resolution in previous works [TL93, TL95b], and it will also been used so in this thesis.

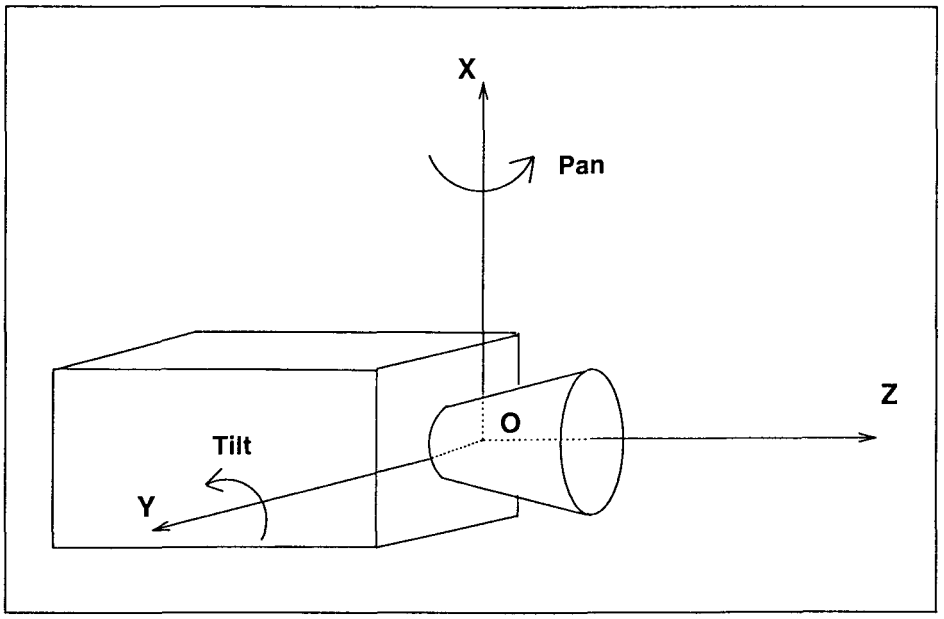


Figure 3.2: Camera pan-tilt model

rotation has two degrees of freedom: rotation about the vertical axis and rotation about the horizontal axis. They are called *pan* and *tilt*, correspondingly. The tilt axis is collinear with the camera Y-axis, which is parallel to the ground. On the other hand, the pan axis is collinear with the camera X-axis, only when the tilt angle is zero. If tilt angle is nonzero, the camera X-axis deviates from the vertical direction. Camera roll (rotation about Z-axis) does not occur. Therefore, the camera Y-axis is always horizontal.

Assume that the pan angle is α and the tilt angle is β. Since the tilt axis is collinear with the camera Y-axis, the transformation matrix for tilt is a standard rotation matrix about Y,

$$T_{tilt}(\beta) = Rot_Y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}. \tag{3.6}$$

Pan is identical to the rotation about X, only when the camera tilt angle is zero. For nonzero tilt angle θ to rotate about the vertical axis, pan is a composite movement of three rotations: First, a rotation about Y with the $-\theta$ angle aligns the X axis to the vertical axis. Second, the camera rotates about X with angle α. Finally, a rotation about Y with angle θ returns the camera back to its tilt position. The transformation matrix can be represented as;

$$T_{pan}(\alpha) = Rot_Y(\theta)Rot_X(\alpha)Rot_Y(-\theta) \tag{3.7}$$

$$= \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}.$$

Camera pan-tilt motions have been represented as transformation matrices. The end position of the camera, or so-called camera pose, can be described by a pair of pan-tilt position angles α and β with respect to the original position. Consequently, the transformation matrix from the original camera coordinate to the new camera coordinate is $Rot_Y(\beta) \cdot Rot_X(\alpha)$, where α, β are the position angles of the camera. This transformation matrix reflects the fact that the Y-axis of the camera coordinate always remains horizontal.

Since a camera-centered coordinate system is used, the world has to be represented in the new camera coordinates. From the discussion above, by relative movement, the world moves in the opposite direction. Therefore, the transformation matrix $T_{world}(\alpha, \beta)$ of the world to new camera position with position angle α and β is $T_{world}(\alpha, \beta) = Rot_Y(-\beta) \cdot Rot_X(-\alpha)$.

$$T_{world}(\alpha,\beta) = \begin{bmatrix} \cos\beta & \sin\alpha\sin\beta & -\cos\alpha\sin\beta \\ 0 & \cos\alpha & \sin\alpha \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix} \tag{3.8}$$

With the transformation matrix given by Eq. (3.8), the projection of the world can be studied under the new camera coordinate. Two types of lines with special characteristics are the vertical and horizontal lines. In the original camera coordinates, the X-axis is vertical and the Y-axis is horizontal. Therefore, direction cosines for the vertical and horizontal lines are $b_v = (1\ 0\ 0)^T$ and $b_h = (0\ 1\ 0)^T$ respectively, in the original camera coordinates. In the new camera coordinates, for which the camera is positioned with pan-tilt angles α and β, the direction cosines can be computed as $b' = T_{world}(\alpha,\beta)\cdot b$. The results are given in Eq. (3.9) where the subscript v applies to a vertical line and the subscript h applies to a horizontal line.

$$b'_v = \begin{bmatrix} \cos\beta \\ 0 \\ \sin\beta \end{bmatrix}, \qquad b'_h = \begin{bmatrix} \sin\alpha\sin\beta \\ \cos\alpha \\ -\sin\alpha\cos\beta \end{bmatrix} \tag{3.9}$$

Eq. (3.9) gives the direction cosines of vertical and horizontal lines in the new camera coordinates. From Eqs. (3.4) and (3.9), the vanishing points of the vertical lines on the image plane are [RGL95]:

$$x_{v\infty} = \frac{f}{\tan\beta}, \qquad y_{v\infty} = 0 \tag{3.10}$$

The vanishing points of the horizontal lines on the image plane are [RGL95]:

$$x_{h\infty} = -f\tan\beta, \qquad y_{h\infty} = \frac{-f}{\tan\alpha\cos\beta} \tag{3.11}$$

It is shown that vanishing points of vertical lines are always on the x-axis of the image coordinate. The location of the vanishing point depends only on the camera tilt position angle. This is due to the particular pan-tilt model being used for which the camera Y-axis remains in the horizontal plane, no matter how the camera rotates. Vanishing points will be used in later sections to develop a method for compensating the perspective distortion.

3.3 Perspective Distortion Compensation by the S-RWT

In Section 3.2, it has been shown that parallel lines with direction cosines $b_3 \neq 0$ converge to vanishing points on the image plane. Generally speaking, the projection is not shape-preserving and is perspective distorted. Recalling the discussion in Section 3.1, the frontal plane is the kind of object plane that is parallel to the image plane. In other words, the coplanar structures on the frontal plane all have the direction cosines $b_3 = 0$. The non-parallelism of the image plane and object plane is the cause of perspective distortion. The camera Z-axis is required to be perpendicular to the object plane in order to get a frontal view.

Unfortunately, a frontal view is not always available. Due to the geometrical structure of the corridor environment and the active camera movement, the camera does not usually point straight to the object plane. The position and orientation of the image plane are dependent on the final camera pose. From an arbitrary viewpoint, the final camera pose with respect to the object plane has six degrees of freedom: three for rotation and three for translation. Due to the high dimension of the search space,

direct model matching for object recognition is not feasible. The frontal view reduces the camera pose to four degree of freedom: one rotation about the camera Z-axis and three translations. The translation along the Z-axis scales the projected image. For example, translation directly towards or away from the frontal plane can cause a uniform change of scale in the projected image. The canonical frontal view further removes the rotation parameter from the frontal view. The process of converting a perspective distorted view to a canonical frontal view is called *perspective distortion compensation.*

In this and the next chapter, a novel approach is developed to compensate for perspective distortion by using the *Shifted Reciprocal Wedge Transform* (*S-RWT*) and *Generalized Shifted Reciprocal Wedge Transform* (*GS-RWT*). This technique is an extension of the *Reciprocal Wedge Transform* (*RWT*) developed by Tong and Li[TL94, TL95a, TL95b]. For road navigation case, the camera always points forward, and, therefore, the image plane is perpendicular to the object plane which is the road plane. Road boundaries are considered as $2D$ features on the object plane. Tong and Li showed that the *RWT* was able to recover the projective distortion with a camera pose that has 90° interception angle between the image plane and the object plane.

In the case of corridor navigation, the camera pose is random due to the pan-tilt movement. The original camera pose is such that camera Z-axis is perpendicular to the object plane and the pan and tilt angles are zero. In such a view, i.e. the canonical frontal view, the object model can be directly applied to the image for matching without rotation. Any camera pan-tilt will cause the loss of the canonical frontal view. As discussed in Section 3.2, the pan-tilt device has two axes of freedom of rotation. In this section, a method based on the *S-RWT* is developed to compensate the distortion

caused by camera rotation in one axis. The discussion assumes a single axis camera rotation(tilt). In Chapter 4, the method is extended to handle the distortion caused by rotation in 2 degrees of freedom by using the *GS-RWT*.

3.3.1 Perspective distortion caused by camera tilt

Two coordinate systems are used: the camera coordinate system and the world coordinate system. The original setting of these two coordinate systems is illustrated in Fig. 3.3. At the original position, the world coordinate can be translated from the camera coordinates in the Z direction. The object plane is set at $Z_{world} = 0$. $2D$ landmarks are represented in world coordinates. At the original position, when the camera pan-tilt angles are zero, the image plane is parallel to the object plane and there is no rotation between the two coordinate systems; the object plane is a canonical frontal plane and the image taken is a canonical frontal view. Once the camera starts to pan-tilt, the camera-centered representation is used. The coordinate with respect to world have to be converted into camera coordinate.

In this section, we assume camera rotation with only one degree of freedom. A random point on the object plane can be represented as $[X, Y, 0]^T$ with respect to the world coordinates. At the original setting, when the pan-tilt angles are zero, the point can be represented as $[X, Y, Z_c]^T$ in the camera coordinates, where Z_c is a translation between the coordinate systems. After camera tilt occurs, the point can be represented in the new camera coordinate. The transformation is given by Eq. (3.8) with pan angle $\alpha = 0$. Therefore, the coordinate after camera tilt with respect to the new camera coordinate is $[X', Y', Z']^T = T_{world}(0, \beta) \cdot [X, Y, Z_c]^T$:

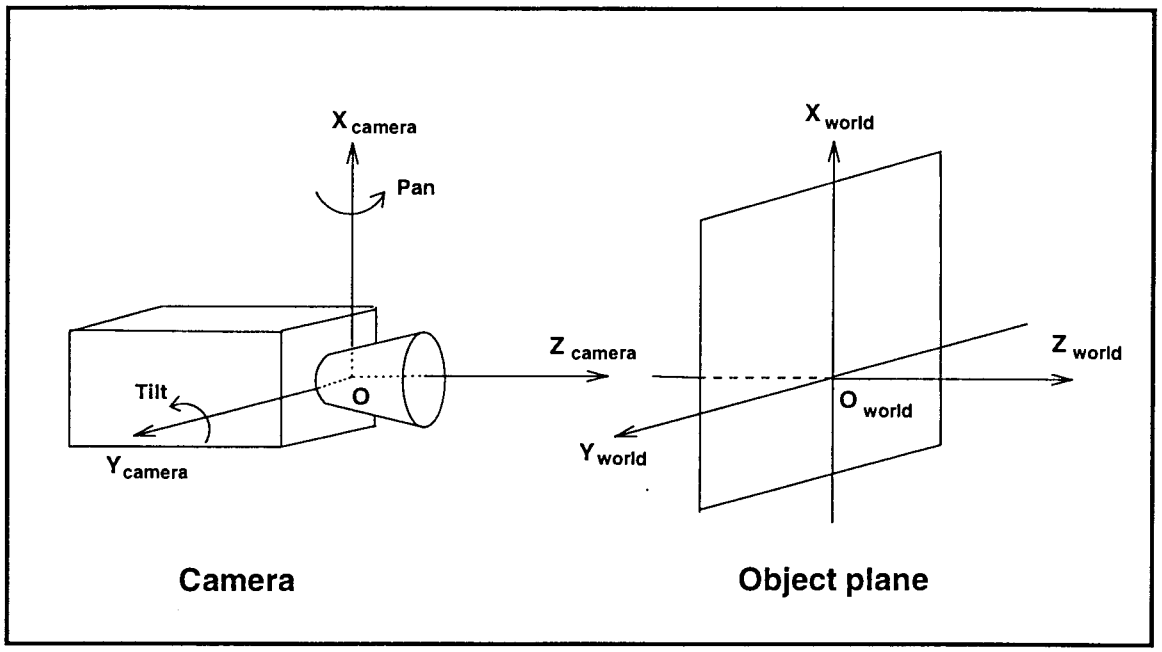


Figure 3.3: Original camera position

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} cos\beta & 0 & -sin\beta \\ 0 & 1 & 0 \\ sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z_c \end{bmatrix} = \begin{bmatrix} Xcos\beta - Z_c sin\beta \\ Y \\ Xsin\beta + Z_c cos\beta \end{bmatrix} \tag{3.12}$$

The projection of $[X', Y', Z']^T$ on the image plane is denoted as $[x_{image}, y_{image}]^T$, which can be calculated by $[x_{image}, y_{image}]^T = [f\frac{X'}{Z'}, f\frac{Y'}{Z'}]^T$, where f is the camera focal length. Therefore,

$$\begin{bmatrix} x_{image} \\ \\ y_{image} \end{bmatrix} = \begin{bmatrix} f\frac{Xcos\beta - Z_c sin\beta}{Xsin\beta + Z_c cos\beta} \\ \\ f\frac{Y}{Xsin\beta + Z_c cos\beta} \end{bmatrix}. \tag{3.13}$$

This is obviously not the canonical frontal view or frontal view. The image is perspective distorted, and there is no simple relations between image pixels and original world pixels. Distortion is caused by the non-parallelism between the image plane

and object plane. Nevertheless, the vanishing points of the parallel lines still contain some useful information.

The vanishing points of the vertical and horizontal lines can be calculated using Eq.s (3.10) and (3.11) with the camera pan angle $\alpha = 0$:

$$x_{v\infty} = \frac{f}{\tan\beta}, \qquad y_{v\infty} = 0 \tag{3.14}$$

$$x_{h\infty} = -f\tan\beta, \qquad y_{h\infty} = -\infty \tag{3.15}$$

Eq. (3.14) and (3.15) show that all the vertical lines will converge to $x_{image} = \frac{f}{\tan\beta}$; and all the horizontal lines will remain horizontal since the vanishing point is at the infinity. This can be shown geometrically in Fig. 3.4. The right figure is the canonical frontal view of a rectangular shape on the object plane. In this case, there is no camera pan or tilt; the camera Z axis is perpendicular to the object plane. Thus, a rectangular shape is also rectangular on the image plane. The left figure is the image obtained by a camera tilted up from the original position. The image is foreshortened in one direction which is the x-axis of the image coordinate. Therefore horizontal edges remains horizontal; and the vertical edge converges to $x = \frac{f}{tan\beta}$, where f is the focal length and β is the tilt angle.

3.3.2 Projective model

In the previous section, it is shown that camera tilt introduces perspective distortion. The most straightforward way to remove perspective distortion is rotating the camera such that the image plane is parallel to the object plane. Consequently, the object plane becomes frontal or canonical frontal. In practice, however this significantly

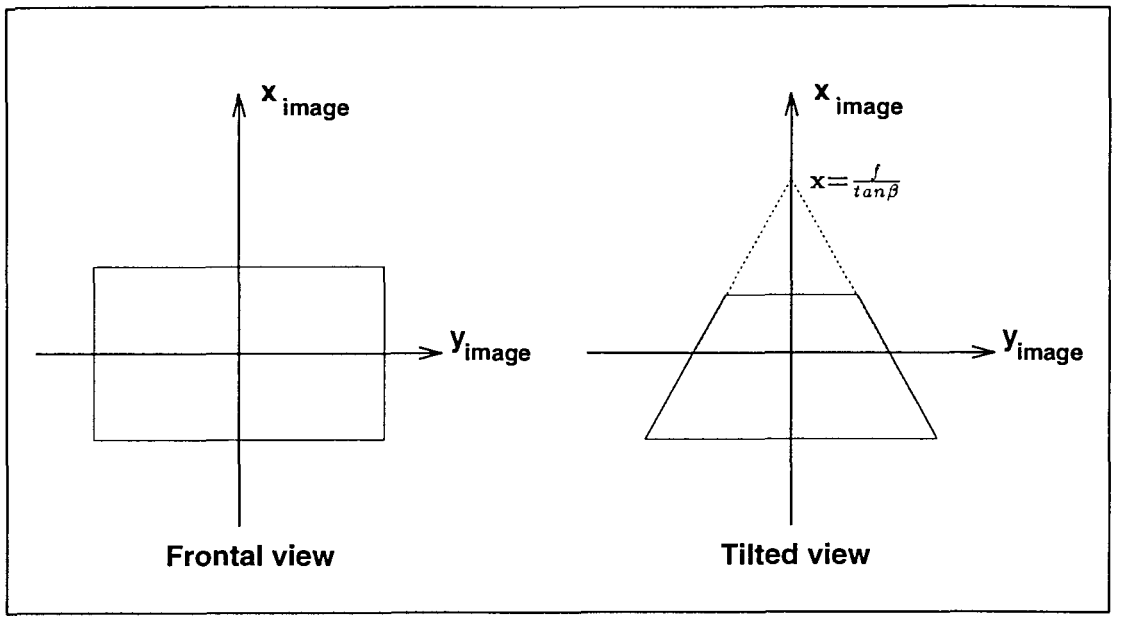


Figure 3.4: Perspective distortion caused by camera tilt

limits the possible motion of the camera. It violates the principles of active vision in the following two ways. First, because of the limitation of the camera lens angle and the size of the image plane, the camera has to be rotated to view the desired objects. Second, from the active vision point of view, camera movement is essential for system intelligence and efficiency. Therefore, limiting the camera pose is not practical.

Some $2D$ transformations of image coordinates can convert a distorted image to its canonical frontal view, which simulates camera rotation. The transformation is not unique [Kan88, Kan93]. When the camera is tilted, there are four degrees of the freedom for camera pose: one for rotation and three for translation. By knowing the camera tilt angle, the transformation should be able to recover the canonical frontal view. As discussed before, the distortion is essentially an image foreshortening in the x direction. It is caused by the non-parallelism between the image plane and the object plane. In other words, the image plane and the object plane form a V-plane

projection [Ton95]. In Chapter 2, the *S-RWT* is proved to have the same effect as V-plane projection. In this section, a projective model is presented to demonstrate the V-plane projection and the transformation parameters for *S-RWT* are derived.

A projective model for camera tilt is illustrated in Fig. 3.5. There are four coordinates involved: camera coordinates, world coordinates, image coordinates and *S-RWT* coordinates. Since a camera-centered view is taken, camera coordinates are used as the reference coordinates. Camera and world coordinates are $3D$ coordinates as shown in Fig. 3.3. Because camera tilt is a rotation about the camera Y-axis, the X, Z axes of the camera and world coordinates are coplanar after rotation. Most of the important features are shown on the X-Z plane. Fig. 3.5 is a $2D$ plot of the X-Z plane. Y-axes of the image and the world coordinates point into the paper and are not shown in this figure. Camera tilt causes an interception angle β between the image X-Y plane and the world X-Y plane. The object plane is set on the world X-Y plane. The translation Z_c between the camera origin and the object plane is not affected by camera tilt.

The image coordinate is on the camera image plane which is parallel to the camera X-Y plane, and intersects the Z_{camera} axis at $Z_{camera} = f$ where f is the camera focal length. Suppose x, y axes are chosen to be the same as those of the camera X, Y axes, and suppose the plane rotates with the camera.

The *S-RWT* coordinates are on an imaginary plane used to re-project the camera image and with principal axes u and v. After rotation, the imaginary *S-RWT* plane is chosen to be parallel to the object plane X_{world}-Y_{world}, and intersects the X_{camera} axis at $X_{camera} = 1$. A unit distance is chosen to simplify the derivation. The origin of the u-v axes is $(1, 0, 0)$ which is measured with respect to the camera coordinates. Such a

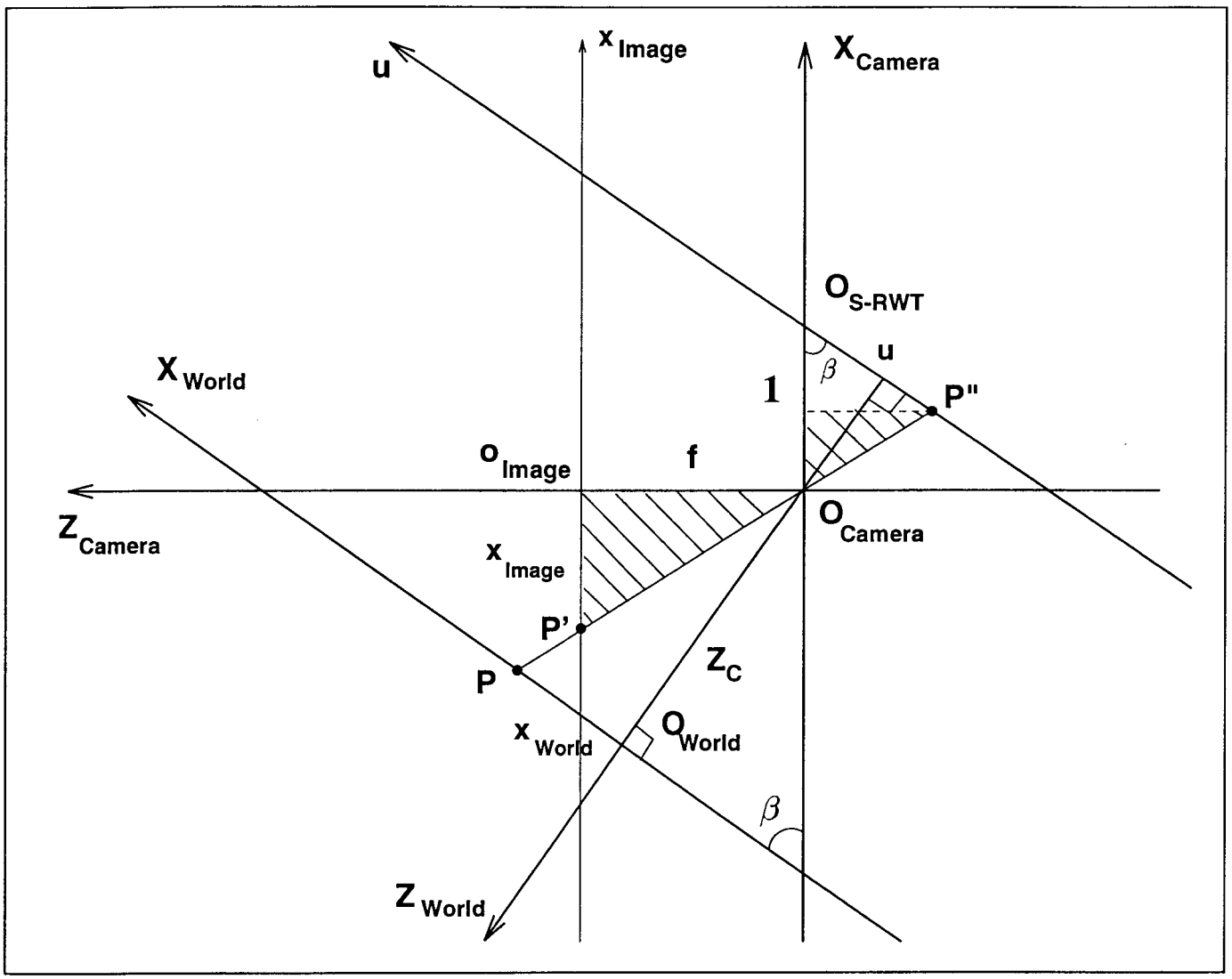


Figure 3.5: Projective model of perspective distortion compensation by the S-RWT (only the X-Z plane is shown)

choice of the origin is convenient for rotations, since u, v axes have the same directions as those of world X, Y axes. Consequently, the v-axis has the same direction as the camera Y-axis.

Suppose the x-y plane and the *S-RWT* u-v plane form a V shape as illustrated in Fig. 3.5. The re-projection of an image pixel onto the *S-RWT* plane through the camera optical center O_{camera} is a V-plane projection[TL95b]. The interception angle between them is β. In the rest of this section, it will be shown that the re-projected image onto the *S-RWT* plane is a canonical frontal view of the object plane, and an *S-RWT* will be established for this V-plane projection.

A random point P on the object plane is projected onto the image plane at P' and re-projected onto the *S-RWT* plane at P''. P is represented as (X, Y, Z_c) in the original camera coordinate. After the camera tilt, the coordinate of P' on the image plane is shown in Eq. (3.13). As shown in Fig. 3.5, the coordinate of P'' on the *S-RWT* plane can be calculated by similar triangles:

$$\frac{X}{Z_c} = \frac{-u - cos\beta}{sin\beta} \tag{3.16}$$

$$\frac{Y}{Z_c} = \frac{-v}{sin\beta} \tag{3.17}$$

where X, Y are the world coordinates, and u, v are the *S-RWT* coordinates.

By moving u and v to the right hand side of Eqs (3.16) and (3.17), the projection point P'' of P on the *S-RWT* plane is,

$$u = -X\frac{sin\beta}{Z_c} - cos\beta \tag{3.18}$$

$$v = -Y\frac{sin\beta}{Z_c} \tag{3.19}$$

From the above result, it can be shown that the re-projected image on the *S-RWT* plane is an canonical frontal view of the object plane. The correspondence point is scaled by $-\frac{sin\beta}{Z_c}$ and translated by $-cos\beta$ in the u direction. Because Z_c is a constant and the tilt angle β is known, the scaling and translational factors are constants. Therefore, perspective distortion has been eliminated. This is an expected result, since the imaginary *S-RWT* plane is parallel to the object plane.

As illustrated above, the image on the *S-RWT* plane is an canonical frontal view of the object plane. The second step will show how to transform the distorted image on the image x-y plane to the *S-RWT* u-v plane. This can be done by obtaining the projective correspondence points P' and P''. For the shaded area in Fig. 3.5, the triangles are similar. By equalizing the length proportions of the sides, we obtain:

$$\frac{-x}{f} = \frac{1 + ucos\beta}{-usin\beta} \tag{3.20}$$

$$\frac{y}{f} = \frac{v}{usin\beta} \tag{3.21}$$

where x, y are relating to the image coordinates and u, v are relative to the *S-RWT* coordinates.

By rearranging Eqs (3.20) and (3.21), the transformation from image plane to *S-RWT* plane is,

$$u = \frac{\frac{f}{sin\beta}}{x - \frac{f}{tan\beta}} \tag{3.22}$$

$$v = \frac{y}{x - \frac{f}{tan\beta}} \tag{3.23}$$

This latter transform is the *Shifted Reciprocal-Wedge Transform*[TL95b] with shift

constant $a = \frac{f}{\tan\beta}$[2]. Since projection of the original camera image to the imaginary *S-RWT* plane forms a V-plane projection with the projection center at the camera optical center (see Fig. 3.5), this relation matches the result from [Ton95]: *S-RWT* can be implemented with V-plane projection. The shift constant depends upon the interception angle of the V-shape. When the camera is tilted, the intersection angle is equal to the tilt angle β.

This result provides an extension of the *RWT* to road navigation [TL94, TL95a, TL95b], for which the camera viewing direction is always parallel to the road (object plane). The interception angle β for V-plane projection is 90°. By Eq. (3.22) and (3.23) , $u = \frac{f}{x}$ and $v = \frac{y}{x}$. Here, we apply the *Reciprocal Wedge Transform*, a special case of *Shifted Reciprocal Wedge Transform* with the shift constant $a = 0$.

From the space-variant image representation point of view, the vanishing point gives a natural explanation for the above transformation. Referring to Fig. 3.4, in the case of camera tilt, the image is foreshortened in the image x direction. Vertical lines on the object plane converge at $x = \frac{f}{\tan\beta}$ by Eq. (3.14). To get an canonical frontal view, converging lines must be remapped to vertical lines in order to preserve the original shape. As discussed in Chapter 2, the *S-RWT* with a shift constant $a = \frac{f}{tan\beta}$ maps the lines concurrent at the x intercept $a = \frac{f}{tan\beta}$ in the x-y domain to u-v lines are vertical. Therefore, Eq. (3.22) and (3.23) compensates the perspective distortion.

Fig. 3.6 summarizes the procedure for perspective distortion compensation. The shift constant is equal to the x coordinate of the vanishing point of the vertical lines. The x coordinate of the vanishing point is dependent on the camera tilt angle, which

[2]A scale factor $\frac{f}{sin\beta}$ in Eq. (3.22) is applied in order to compensate the perspective distortion.

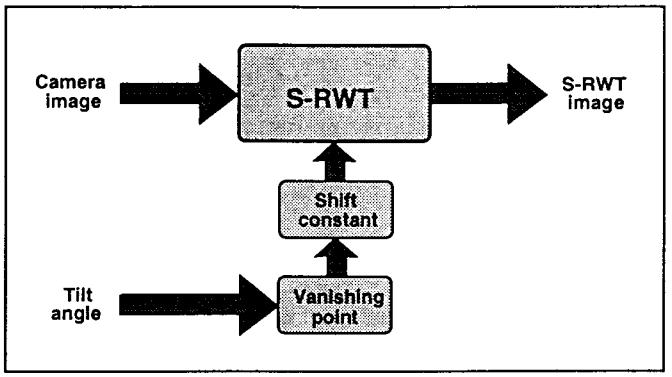


Figure 3.6: Perspective distortion compensation by S-RWT

is also the interception angle for the V-plane projection formed by the image plane and the *S-RWT* plane. Applying the *S-RWT* to the distorted image with a shift constant compensates the distortion caused by camera tilt.

3.3.3 Perspective distortion compensation by S-RWT

In the previous section, a projective model for V-plane projection was presented to obtain a canonical frontal view of the object plane. The interception angle of the V-plane projection is the camera tilt angle β. The V-plane projection was represented by a *S-RWT* with shift constant $a = \frac{f}{\tan\beta}$. The inspiration of this approach comes from the following considerations:

1. The *S-RWT* has the characteristic that lines exchange the slope and intercept in the x-y domain and the u-v domain.
2. The *S-RWT* can be implemented by a V-plane projection.

In this section, the whole process is put together to prove the effectiveness of perspective distortion compensation when the camera tilted. The *S-RWT* is applied

to the distorted image and we show that the result image forms an canonical frontal view.

Using a camera-centered coordinate system, a random point on the object plane before camera tilt is $[X, Y, Z_c]$, where Z_c is a constant. After the camera tilt, the point is represented in new camera coordinates, and projected onto the image plane. By Eq. (3.13), the image coordinate for the projected point is,

$$\begin{bmatrix} x_{image} \\ \\ y_{image} \end{bmatrix} = \begin{bmatrix} f\frac{X cos\beta - Z_c sin\beta}{X sin\beta + Z_c cos\beta} \\ \\ f\frac{Y}{X sin\beta + Z_c cos\beta} \end{bmatrix}. \tag{3.24}$$

Eq. (3.24) represents the distorted image on the camera image plane. By Eq. (3.22), (3.23) and Eq. (3.24), the *S-RWT* is applied to the distorted image with the shift constant $a = \frac{f}{tan\beta}$:

$$\begin{aligned} u &= \frac{\frac{f}{sin\beta}}{x_{image} - \frac{f}{tan\beta}} \\ &= \frac{\frac{f}{sin\beta}}{f\frac{X cos\beta - Z_c sin\beta}{X sin\beta + Z_c cos\beta} - \frac{f}{tan\beta}} \\ &= -\frac{X}{Z_c} sin\beta - cos\beta \end{aligned}$$

$$\begin{aligned} v &= \frac{y_{image}}{x_{image} - \frac{f}{tan\beta}} \\ &= \frac{f\frac{Y}{X sin\beta + Z_c cos\beta}}{f\frac{X cos\beta - Z_c sin\beta}{X sin\beta + Z_c cos\beta} - \frac{f}{tan\beta}} \end{aligned}$$

$$= -\frac{Y}{Z_c} sin\beta$$

The result is the same as that derived from the projective model. It differs from the original point on the object plane by a scaling factor $-\frac{\sin\beta}{Z_c}$ and a translational parameter $-\cos\beta$ in the u direction. The image after the *S-RWT* is an canonical frontal view of the object plane. Nevertheless, this *S-RWT* based method has the limitation that it can only handle the distortion caused by camera tilt. In next chapter, a new technique called the *Generalized Shifted Reciprocal-Wedge* Transform will be introduced to compensate the perspective distortion caused by both camera pan and tilt.

Chapter 4

Distortion Compensation by the GS-RWT

In Chapter 3, a method based on the *S-RWT* was developed to compensate the perspective distortion caused by a camera tilt. After the camera tilt, the image is foreshortened in the image x direction. The distortion can be compensated by applying the *S-RWT* with a shift constant a, calculated from the tilt angle β such that $a = \frac{f}{\tan\beta}$.

In this chapter, this approach is extended to handle the distortion caused by both the camera pan and tilt. For a point $[X, Y, Z_c]^T$ on the object plane, by Eq. (3.8), the coordinate after camera pan-tilt is $[X', Y', Z']^T = T_{world}(\alpha, \beta) \cdot [X, Y, Z_c]^T$ where

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\alpha\sin\beta & -\cos\alpha\sin\beta \\ 0 & \cos\alpha & \sin\alpha \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z_c \end{bmatrix}$$

$$= \begin{bmatrix} X cos\beta + Y sin\alpha - Z_c cos\alpha sin\beta \\ Y cos\alpha + Z_c sin\alpha \\ X sin\beta - Y sin\alpha cos\beta + Z_c \cos\alpha cos\beta \end{bmatrix} \tag{4.1}$$

and Z_c is a constant.

The projection of $[X', Y', Z']$ on the image plane is denoted by $[x_{image}, y_{image}]^T$, which can be calculated by $[x_{image}, y_{image}]^T = [f\frac{X'}{Z'}, f\frac{Y'}{Z'}]^T$. Therefore

$$\begin{bmatrix} x_{image} \\ \\ y_{image} \end{bmatrix} = \begin{bmatrix} f\frac{X cos\beta + Y sin\alpha - Z_c cos\alpha sin\beta}{X sin\beta - Y sin\alpha cos\beta + Z_c cos\alpha cos\beta} \\ \\ f\frac{Y cos\alpha + Z_c sin\alpha}{X sin\beta - Y sin\alpha cos\beta + Z_c cos\alpha cos\beta} \end{bmatrix}. \tag{4.2}$$

Eq. (4.2) represents the distortion of the image. We assume that the camera pose has five degrees of freedom: two for rotation and three for translation. The *Generalized Shifted Reciprocal Wedge Transform* (*GS-RWT*) is introduced to eliminate the two rotational parameters from the projected image. The resulting image will be a canonical view of the object plane.

The *GS-RWT* is a $2D$ transform performed on the image plane. It can be decomposed into two $2D$ rotations and one *S-RWT*. Therefore, the *GS-RWT* is parameterized by two rotational angles and a shift constant. The *GS-RWT* maintains the properties of the *RWT* and the *S-RWT*: space-variant sensing and linearity preservation. Moreover, the direction of space-variant sensing can be adjusted by varying the angle of the first rotation.

4.1 Camera pan-tilt geometry

Perspective distortion is caused by the non-parallelism of the image plane and object plane. In this section, the geometrical configurations of the image plane and object

plane after camera pan-tilt are studied in order to give the foundation for the *GS-RWT*. As shown in Fig. 3.3, at the original setting, the object plane is a canonical plane and the camera image is a canonical view. After camera pan-tilt, the object plane and camera image plane are illustrated in Fig. 4.1. Since the image plane is not parallel to the object plane, the projection of the object on the image plane is perspectively distorted. As represented in Eq. (4.2), some terminology introduced in Fig. 4.1, is defined below,

X_{world}, Y_{world}:	$2D$ coordinate of object plane.
x_{image},y_{image}:	$2D$ coordinate of image plane.
L_y:	interception vector of object plane and image plane.
L_x:	vector perpendicular to L_y and parallel to image plane.
L'_x:	vector perpendicular to L_y and parallel to object plane.
δ:	interception angle of L_x and x_{image}.
γ:	interception angle of L'_x and X_{world}.
θ:	interception angle of L_x and L'_x.

Since L_y, L_x and L'_x are vectors (the position values of these parameters are not important for *GS-RWT*), L_y can be computed from the interception of the image plane and the object plane. L_x is perpendicular to L_y and parallel to the image plane. L'_x is perpendicular to L_y and parallel to the object plane. Consequently, the interception angle θ of L_x and L'_x is also the interception angle of the image and the object plane.

As shown in Fig. 4.1, if the camera pan or tilt angle is not zero, the image plane intersects the object plane. For a camera tilt with a tilt angle β, the intersect is a horizontal line. Therefore vector L_y is parallel both to y_{image} and Y_{world}, L_x is parallel

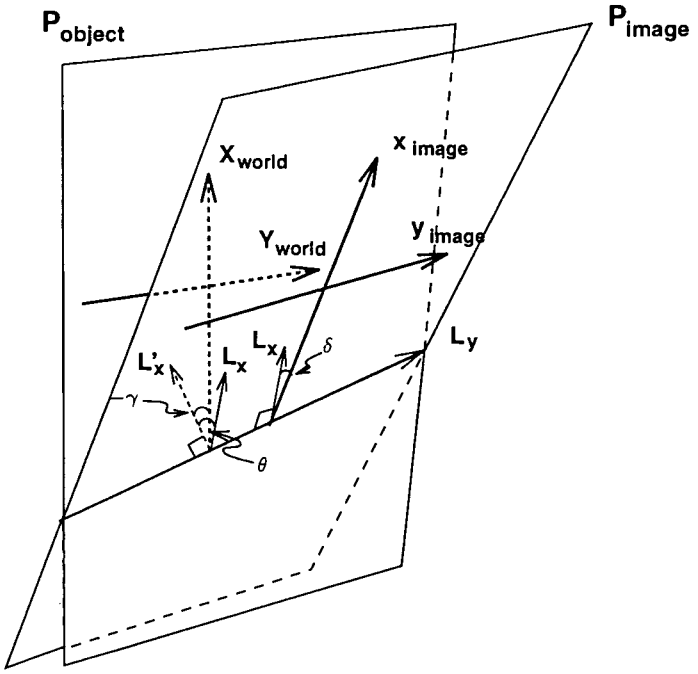


Figure 4.1: Configuration of image plane and object plane after camera pan-tilt

to x_{image}, and L'_x is parallel to X_{world}. Both interception angles δ and γ are zero. The interception angle θ of L_x and L'_x is equal to camera tilt angle β.

V-plane projection foreshortens the image. When the camera is tilted, δ is zero, and the direction of the foreshortening is along the x-axis in the image coordinate. The interception angle θ of the V shape determines the rate of foreshortening. Since the vanishing point is at $x = \frac{f}{\tan\beta}$ (see Fig. 3.4), the greater the β, the closer the vanishing point is to the origin. The resolution of the image decreases along x towards the vanishing point. Therefore, the greater the β, the more significant is the foreshortening. For a canonical frontal view with zero pan-tilt angles, the vanishing point becomes infinity so that no foreshortening exists.

For non zero pan angle α and tilt angle β, the image plane and object plane have a configuration similar to that obtained for only camera tilt, except the interception

vector L_y is not parallel to y_{image} or Y_{world}. Consequently, L_x is not parallel to x_{image} and L'_x is not parallel to X_{world}. The interception angles δ and γ are nonzero. The angle θ between L_x and L'_x depends upon the pan-tilt angles α and β.

The image plane and object plane still form a V-plane projection with angle θ. Compared to the camera tilt case, the V shape is tilted such that the "valley" has the direction of the current L_y instead of the horizon. It can be imagined that the projection from the object plane to image plane will still have the effect of foreshortening, regardless of the tilt of the V shape. The direction of the foreshortening is along L_x on the image plane. But, since δ is nonzero, L_x is not x_{image}. Therefore, the resolution of the image pixels varies along L_x instead of x_{image}, and all pixels on the same parallel line of L_y have a unique resolution.

To compensate for the distortion of foreshortening along the direction L_x on the image plane, the *S-RWT* is not sufficient, since *S-RWT* only foreshortens the image in the x direction. In addition we must perform a rotation before the *S-RWT* to align the L_x to the x-axis. The amount of rotation is δ the interception angle between L_x and the x-axis. After rotation, the *S-RWT* can be applied. The shift constant varies with the rate of foreshortening which is determined by the interception angle θ of the V shape. As a hypothesis, shift constant could have the same form as that for camera tilt, $a = \frac{f}{\tan\theta}$.

As discussed previously, application of a rotation and the *S-RWT* might compensate the distortion. Assuming that all the hypotheses are correct, the image after rotation and the *S-RWT* should be a frontal view, but not yet a canonical frontal view. With the camera tilt, the camera only rotates about Y_{camera}-axis, x_{image} and

X_{world} are coplanar, and L'_x is parallel to X_{world}. Therefore the image after the *S-RWT* is already canonical frontal. With camera pan-tilt, L'_x is not parallel to X_{world} and they have an interception angle γ. To achieve a canonical front view, L'_x has to be rotated back to X_{world} with angle $-\gamma$.

From the above discussion, we have analyzed the geometrical configurations of the image plane and the object plane after camera pan-tilt. Analogous to the case of camera tilt, we will develop an approach to compensate the perspective distortion caused by pan-tilt called the *Generalized Shifted Reciprocal Transform.* For the analysis in this section, the following hypothesis is made:

Hypothesis: the *GS-RWT* can be used to compensate for distortion caused by camera pan-tilt in the following three steps:

1. Rotate the image by the angle δ about the origin .
2. Apply the *S-RWT* with shift constant $a = \frac{f}{\tan\theta}$.
3. Rotate the resulting image by an angle $-\gamma$ about the origin.

The parameters δ, θ and γ are the same as in Fig. 4.1.

4.2 Transformation of the GS-RWT

The *Generalized Shifted Reciprocal-Wedge Transform* is an extension of the *S-RWT*. The *GS-RWT* is a $2D$ transform performed on the image domain. It can be represented as a composition of a rotation in the x-y domain, an *S-RWT* which maps the image pixel on x-y domain to a u-v domain, and a rotation in the u-v domain,

$$T_{GS-RWT}(\gamma, a, \delta) = Rot(-\gamma)T_{S-RWT}(a)Rot(\delta) \tag{4.3}$$

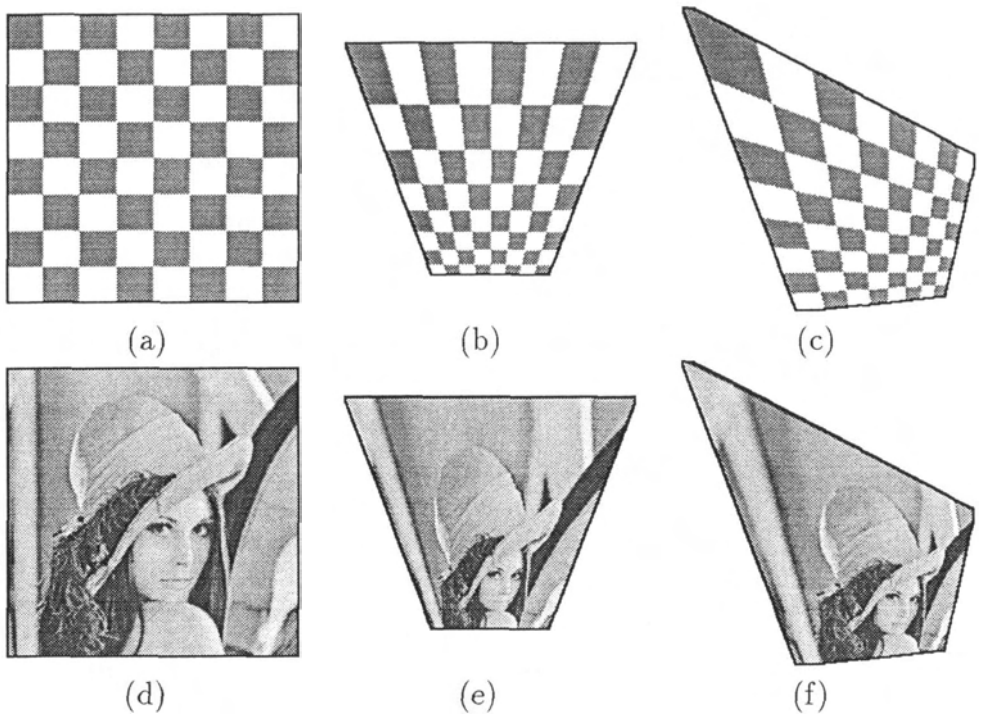


Figure 4.2: The Generalized Shifted Reciprocal Wedge Transform
(a) A rectangular grid. (b) The *S-RWT* image which foreshortens the image in vertical x-axis. (c) The *GS-RWT* image in which the direction of foreshortening deviates from principal axes. (d) The lady's image. (e) *S-RWT* image of the lady's image. (f) *GS-RWT* image of the lady's image.

$$= \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & a \end{bmatrix} \begin{bmatrix} \cos\delta & -\sin\delta & 0 \\ \sin\delta & \cos\delta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sin\gamma\sin\delta & \sin\gamma\cos\delta & \cos\gamma \\ \cos\gamma\sin\delta & \cos\gamma\cos\delta & -\sin\gamma \\ \cos\delta & -\sin\delta & a \end{bmatrix}$$

where γ and δ are rotation angles and a is the shift constant.

As discussed in Chapter 2, the *S-RWT* transforms rectangular images into wedge shapes with variable resolution along the x-axis. The resolution of the image decreases

in the x direction. This is shown in Fig. 4.2 (b) and (e). Vertical direction is the x-axis for x-y image. The shift constant a determines the slope of the wedge edge which is related to the rate of image foreshortening. When apply to perspective distortion compensation, a depends on the interception angle of the V-plane projection.

The *GS-RWT* keeps all the properties of *S-RWT*. Furthermore, with a rotation δ in x-y domain, the *GS-RWT* can foreshorten the image along that direction defined by $x \sin\delta - y \cos\delta = 0$. This feature is shown in Fig. 4.2 (c) and (f) in which the direction of foreshortening deviates from principal axes. This relaxes the restriction that the *S-RWT* provides foreshortening only along the x-axis. A second rotation with rotation angle γ is performed in the u-v domain to get the canonical frontal view.

Since the *GS-RWT* can be decomposed into three single transforms, and each transform has a matrix representation for their inverses, the inverse transform of the *GS-RWT* can be represented by,

$$T_{GS-RWT}(\gamma, a, \delta)^{-1} = Rot(-\delta) T_{S-RWT}(a)^{-1} Rot(\gamma) \tag{4.4}$$

where

$$T_{S-RWT}(a)^{-1} = \begin{bmatrix} -a & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

4.3 Preservation of linearity by the GS-RWT

The *S-RWT* preserves linearity in the x-y domain. The rotations result in orthogonal transforms that are shape-preserving. Since the *GS-RWT* is a composition of *S-RWT* and rotations, the *GS-RWT* also preserves linearity. Using slope and intercept as parameters, a line can be represented as a point in parameter space. The line in

x-y space and u-v is then converted to the mapping of two points in the parameter space. The parameter space is known as the Hough space which gets its name from Hough transform [DH72] which is a powerful tool for line detection. Thus,

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \simeq \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4.5}$$

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

Since a linear transformation exists between the x-y domain and the u-v domain as shown in Eq. (4.5), a line $v = m_{u,v} \cdot u + c_{u,v}$ in the u-v domain is mapped into the x-y domain as:

$$y = \frac{m_{u,v}r_{11} + c_{u,v}r_{31} - r_{21}}{m_{u,v}r_{12} - c_{u,v}r_{32} + r_{22}}x + \frac{m_{u,v}r_{13} + c_{u,v}r_{33} - r_{23}}{-m_{u,v}r_{12} - c_{u,v}r_{32} + r_{22}}$$

The slope of the corresponding line in the x-y domain is $m_{x,y} = \frac{m_{u,v}r_{11}+c_{u,v}r_{31}-r_{21}}{m_{u,v}r_{12}-c_{u,v}r_{32}+r_{22}}$ and its intercept is $c_{x,y} = \frac{m_{u,v}r_{13}+c_{u,v}r_{33}-r_{23}}{-m_{u,v}r_{12}-c_{u,v}r_{32}+r_{22}}$. Therefore, the mapping of the parameters slope and intercept in Hough space is,

$$\begin{bmatrix} m_{x,y} \\ c_{x,y} \\ 1 \end{bmatrix} \simeq \begin{bmatrix} r_{11} & r_{31} & r_{21} \\ r_{13} & r_{33} & -r_{23} \\ -r_{12} & -r_{32} & r_{22} \end{bmatrix} \begin{bmatrix} m_{u,v} \\ c_{u,v} \\ 1 \end{bmatrix} \tag{4.6}$$

From Eq. (4.3) and Eq. (4.6) , the Hough mapping for *GS-RWT* is:

$$M_{GS-RWT} = \begin{bmatrix} \sin\gamma\sin\delta & \cos\delta & \cos\gamma\sin\delta \\ \cos\gamma & a & \sin\gamma \\ -\sin\gamma\cos\delta & \sin\delta & \cos\gamma\cos\delta \end{bmatrix} \tag{4.7}$$

For the special case of *GS-RWT*, where $\gamma = \delta = a = 0$, the properties holds:

$$\begin{bmatrix} m_{x,y} \\ c_{x,y} \\ 1 \end{bmatrix} = \begin{bmatrix} c_{u,v} \\ m_{u,v} \\ 1 \end{bmatrix} \simeq \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{u,v} \\ c_{u,v} \\ 1 \end{bmatrix} \tag{4.8}$$

The slope and intercept parameters of a line in the x-y domain and u-v domains exchange. They form an interesting duality about the main diagonal of the Hough space [TL95b] (See Fig. 4.3).

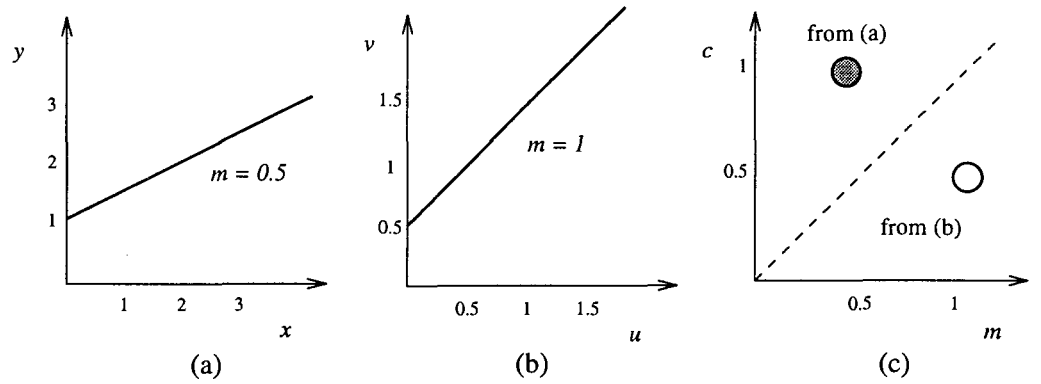
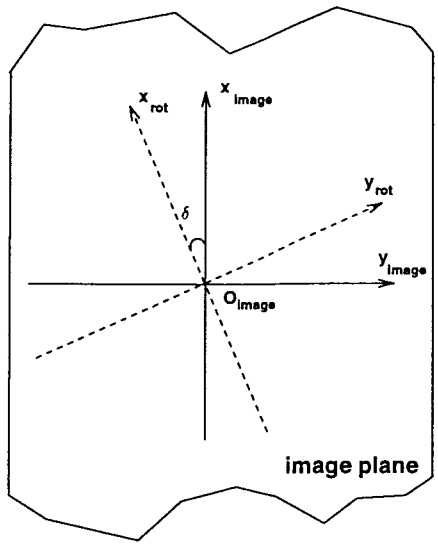


Figure 4.3: Duality of linear structures in the RWT.
(a) A line in the x-y domain with a slope 0.5 and the intercept 1. (b) Dual in the u-v domain, with slope is 1 and intercept 0.5. (c) Hough space showing the peaks from (a) and (b) respectively, reflected about the main diagonal.

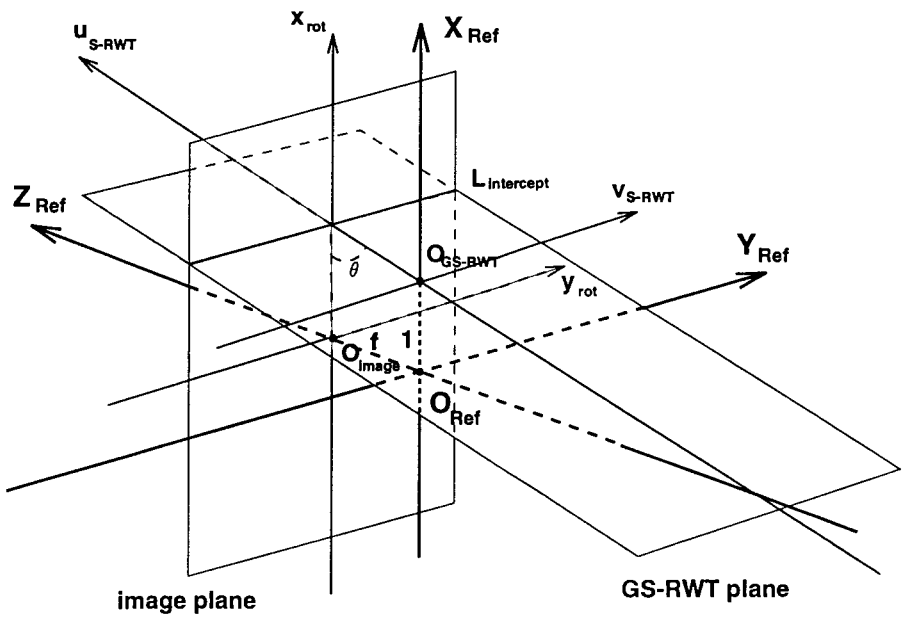
4.4 Projective model of the GS-RWT

Fig. 4.4 illustrates the projective model for the *GS-RWT*. The *GS-RWT* is a composite transform which consists of a rotation, a *S-RWT* and a rotation.(Fig. 4.4 (a), (b) and (c)).

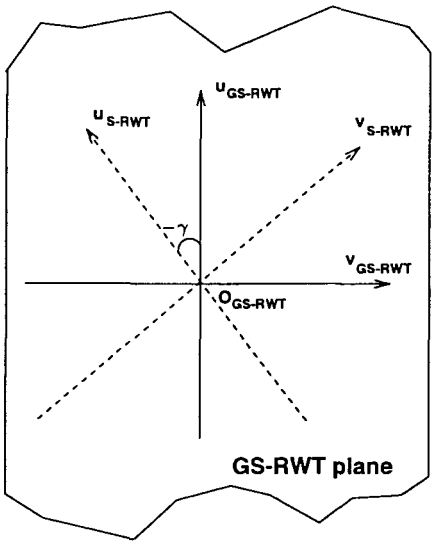
The planes involved in the *GS-RWT* illustrated in Fig. 4.4 (b) are the camera image plane and an (imaginary) *GS-RWT* plane. The image plane is the same as that in Fig. 4.1. Due to camera pan-tilt, the image on image plane is distorted. It


x rot
x image
y rot
δ
y image
O image
image plane

(a)


u S-RWT
x rot
X Ref
Z Ref
L Intercept
v S-RWT
Y Ref
θ
O GS-RWT
y rot
f
1
O image
O Ref
image plane
GS-RWT plane

(b)



(c)

Figure 4.4: GS-RWT projective model.
(a) Original image is rotated by δ in order to foreshorten the image along line $x \sin\delta + y \cos\delta = 0$. (b) *S-RWT* is performed by re-projecting image from image plane to *GS-RWT* plane. (c) A rotation with rotation angle γ is applied on *GS-RWT* plane.

must be projected onto the *GS-RWT* plane to retrieve the canonical frontal view. Therefore, the image plane is the source plane of the *GS-RWT* and the *GS-RWT* plane is the destination plane.

There are two $2D$ coordinate systems on the image plane and they share the same origin, as shown Fig. 4.4 (a). Coordinate system x_{image}-y_{image}, shown in Fig. 4.1, is used to represent the original camera image. The other coordinate system x_{rot}-y_{rot} is generated by rotating x_{image}-y_{image} by a an angle δ. Recall that L_x is perpendicular to L_y on the image plane, the x_{rot}-axis is parallel to L_x and y_{rot} is parallel to L_y. The rotation angle δ is equal to the interception angle between L_x and x_{image}. The transformation from x_{image}-y_{image} to x_{rot}-y_{rot} is represented by $Rot(\delta)$.

The *GS-RWT* plane is an imaginary plane that is always parallel to the object

plane. This requirement is essential for forming an canonical frontal view. The *GS-RWT* plane has reference coordinate system $X_{Ref}Y_{Ref}Z_{Ref}$ as shown in Fig. 4.4 (b). The reference coordinate is obtained when Z_{Ref} is normal to the image plane and it intersects the origin o_{image} of the image coordinates. The origin O_{Ref} is on Z_{Ref} at a distance f from o_{image}, where f is the camera focal length. The X_{Ref}-axis intersects O_{Ref} and is parallel to x_{rot}. The Y_{Ref}-axis intersects O_{Ref} and is parallel to y_{rot}. The reference coordinates can be generated from the camera coordinates by a rotation about Z_{camera} with angle δ. The *GS-RWT* plane intersects X_{Ref} at $X_{Ref} = 1$, and intersects the image plane with an angle θ. The interception line $L_{intercept}$ is parallel to y_{rot} and Y_{Ref}, and, therefore, is parallel to L_x. Referring to Fig. 4.1, the *GS-RWT* plane is parallel to the object plane, because the object plane and the *GS-RWT* plane both have an interception angle θ with the image plane and their interception lines are parallel to L_y.

There are two $2D$ coordinate systems on the *GS-RWT* plane: u_{S-RWT}-v_{S-RWT} and u_{GS-RWT}-v_{GS-RWT}; and they share the same origin o_{GS-RWT}. The origin o_{GS-RWT} is set to $(1, 0, 0)$ with respect to the reference coordinates where the *GS-RWT* plane intersects X_{Ref}. The principal axes u_{S-RWT} and v_{S-RWT} of u_{S-RWT}-v_{S-RWT} intersect the origin o_{GS-RWT}. v_{S-RWT} is set to be parallel to Y_{Ref} and u_{S-RWT} is perpendicular to v_{S-RWT}. With this setting, v_{S-RWT} is also parallel to L_y; and u_{S-RWT} is parallel to L'_x. By comparison with Fig. 3.5, Fig. 4.4 (b) is actually a $3D$ representation of the *S-RWT* projective model. In Fig. 4.4 (b), image plane and *GS-RWT* plane form a V-plane projection with the projection center at O_{Ref} and interception angle θ. Since the axes y_{rot} and v_{S-RWT} are parallel to each other, the mapping from x_{rot}-y_{rot} coordinates to u_{S-RWT}-v_{S-RWT} can be represented as an *S-RWT* with shift constant

$a = \frac{f}{\tan\theta}$. This transformation is the second factor of the *GS-RWT* matrix in Eq. (4.3).

The other coordinate system on the *GS-RWT* plane is u_{GS-RWT}-v_{GS-RWT} which can be generated by a rotation from u_{S-RWT}-v_{S-RWT} about o_{GS-RWT} with an angle $-\gamma$, as shown in Fig. 4.4 (c). The *GS-RWT* plane is parallel to the object plane, as mentioned above. Referring to Fig. 4.1, u_{GS-RWT} can be set to be parallel to X_{world}. Since v_{S-RWT} is parallel to L_y and u_{S-RWT} is parallel to L'_x, the rotation angle $-\gamma$ is equal to the interception angle of L'_x and X_{world}. The coordinate transformation from u_{S-RWT}-v_{S-RWT} to u_{GS-RWT}-v_{GS-RWT} can be represented as the third term of the *GS-RWT* matrix. This rotation is used to convert a frontal view to an canonical frontal view.

The projective model of *GS-RWT* is essentially a projection of the distorted image from the image plane to the *GS-RWT* plane which is parallel to object plane. Therefore, a frontal view of the object plane can be obtained. By rotations on the image plane and *GS-RWT* plane, we expect that an canonical view can be achieved.

4.5 Parameters of the GS-RWT

According to the hypothesis made in Section 4.1, perspective distortion caused by camera pan-tilt can be compensated by the *GS-RWT* as shown in Eq. (4.9) with shift constant $a = \frac{f}{\tan\theta}$. The parameters δ, θ and γ have the same values as those in Fig. 4.1 in order to obtain an canonical frontal view of the object plane. The transformation is

$$T_{GS-RWT}(\gamma, a, \delta) = Rot(-\gamma)T_{S-RWT}(a)Rot(\delta)$$

$$= \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & a \end{bmatrix} \begin{bmatrix} \cos\delta & -\sin\delta & 0 \\ \sin\delta & \cos\delta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.9}$$

The key to the *GS-RWT* is the vector L_y which is determined by the intersection of the image and object planes. The intersection can be calculated as the cross product of image and object plane normal vectors. At the initial position, when the camera pan-tilt angles are both zero, in the camera-centered coordinate system, the normal vectors N_{image} of image plane and N_{world} of object plane are

$$N_{image} = N_{world} = [0 \;\; 0 \;\; 1]^T. \tag{4.10}$$

Because, both the image plane and the object plane are perpendicular to the camera Z-axis. After the camera pan-tilt, the normal vector N_{world} of object plane has to be represented in the new camera coordinates,

$$N'_{world} = T_{world}(\alpha, \beta) \cdot N_{world} = \begin{bmatrix} -\cos\alpha\sin\beta \\ \sin\alpha \\ \cos\alpha\cos\beta \end{bmatrix} \tag{4.11}$$

N'_{world} is the normal in the new camera coordinate, where $T_{world}(\alpha, \beta)$ is the world transformation matrix in Eq. (3.8), and α and β are the camera pan and tilt angles, accordingly. Since the camera coordinate system is fixed as a reference coordinate,

$$N'_{image} = N_{image} = [0 \;\; 0 \;\; 1]^T. \tag{4.12}$$

Therefore, L_y is calculated as the cross product of N'_{world} and N'_{image}. The order of the cross product is such that if α is zero and tilt angle β is positive, L_y has the same direction as y_{image}. Thus,

$$L_y = N'_{world} \times N'_{image} = \begin{bmatrix} \sin\alpha \\ \cos\alpha \sin\beta \\ 0 \end{bmatrix}. \tag{4.13}$$

We observe that the Z component of vector L_y is zero which indicates that L_y is parallel to the X_{camera}-Y_{camera} plane and parallel to the image plane. The vector L_x is defined to be perpendicular to L_y and parallel to the image plane. Therefore, from Eq. (4.13),

$$L_x = \begin{bmatrix} \cos\alpha \sin\beta \\ -\sin\alpha \\ 0 \end{bmatrix}. \tag{4.14}$$

Since L_x is parallel to image plane, the vector representation implicitly indicates the interception angle δ with x_{image}. The rotation matrix with angle δ can be derived from the vector form of L_x directly as we will show.

The shift constant a is assumed to be $\frac{f}{\tan\theta}$, where θ is the interception angle of image plane and object plane. $\cos\theta$ can be calculated from the normals of two intersecting planes. With the normal of the image plane and object plane given by Eq. (4.12) and Eq. (4.11),

$$\cos\theta = \frac{N'_{world} \cdot N'_{image}}{\mid N'_{world} \mid\mid N'image \mid} = \cos\alpha\cos\beta. \tag{4.15}$$

By Eq. (4.15), the shift constant a is

$$a = \frac{f}{\tan\theta} = \frac{f\cos\alpha\cos\beta}{\sqrt{1-\cos^2\alpha\cos^2\beta}} = \frac{f\cos\alpha\cos\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}}. \tag{4.16}$$

The rotation angle γ is the interception angle of L'_x and X_{wall}. Since L'_x is parallel to the object plane and perpendicular to L_y, γ is also a interception angle of L_y and

Y_{wall}, as seen from Fig. 4.1. Therefore, γ can be calculated from the interception angle of L_y and Y_{wall}. The initial vector of Y_{wall} is $[0\ \ 1\ \ 0]^T$, when the camera pan-tilt angles are zero. After camera pan-tilt, Y_{wall} has to be represented in the new camera coordinate system,

$$Y'_{wall} = T_{world}(\alpha, \beta) \cdot Y_{wall} = \begin{bmatrix} \sin\alpha \sin\beta \\ \cos\alpha \\ -\sin\alpha \cos\beta \end{bmatrix}. \tag{4.17}$$

By Eq. (4.13) and Eq. (4.17),

$$\cos\gamma = \frac{Y'_{wall} \cdot L_y}{\mid Y'_{wall} \mid\mid L_y \mid} = \frac{\sin\beta}{\sqrt{\sin^2\alpha + \cos^2\alpha \sin^2\beta}} \tag{4.18}$$

Since the parameters of the *GS-RWT* have been obtained in Eq. 4.14, 4.16 and 4.18, the matrix representation of the *GS-RWT* can be evaluated. By Eq. (4.14), to rotate the x_{image}-axis to L_x, the normalized rotation matrix is,

$$Rot(\delta) = \begin{bmatrix} \frac{\cos\alpha \sin\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha \sin^2\beta}} & \frac{-\sin\alpha}{\sqrt{\sin^2\alpha+\cos^2\alpha \sin^2\beta}} & 0 \\ \frac{\sin\alpha}{\sqrt{\sin^2\alpha+\cos^2\alpha \sin^2\beta}} & \frac{\cos\alpha \sin\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha \sin^2\beta}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.19}$$

A matrix representation for *S-RWT* with shift constant given by Eq. (4.16) can be written as

$$T_{S-RWT}(a) = \begin{bmatrix} 0 & 0 & \frac{f}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} \\ 0 & 1 & 0 \\ 1 & 0 & \frac{-f\cos\alpha\cos\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} \end{bmatrix} \tag{4.20}$$

The term on the upper-right corner of the matrix is a scaling factor on u direction, which is carried out by inspection.

With $\cos\gamma$ given by Eq. (4.18), the normalized rotation matrix $Rot(-\gamma)$ is

$$Rot(-\gamma) = \begin{bmatrix} \frac{\sin\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & \frac{\sin\alpha\cos\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & 0 \\ \frac{-\sin\alpha\cos\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & \frac{\sin\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.21}$$

By Eqs. (4.19), (4.20) and (4.21), the transformation matrix for $GS\text{-}RWT$ is:

$$T_{GS-RWT}(\alpha,\beta) = Rot(-\gamma)\cdot T_{S-RWT}(a)\cdot Rot(\delta) \tag{4.22}$$

$$= \begin{bmatrix} \frac{\sin^2\alpha\cos\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} & \frac{sin\alpha\cos\alpha\sin\beta\cos\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} & \frac{f\sin\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} \\ \frac{\sin\alpha\sin\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} & \frac{\cos\alpha\sin^2\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} & \frac{-f\sin\alpha\cos\beta}{\sin^2\alpha+\cos^2\alpha\sin^2\beta} \\ \frac{\cos\alpha\sin\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & \frac{-\sin\alpha}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} & \frac{-f\cos\alpha\cos\beta}{\sqrt{\sin^2\alpha+\cos^2\alpha\sin^2\beta}} \end{bmatrix}$$

The final matrix representation for $GS\text{-}RWT$ in Eq. (4.22), is generated by combining the three separate transforms. The resulting matrix has the camera pan-tilt angles as dependent variables. The procedure of perspective distortion compensation by the $GS\text{-}RWT$ is summarized in Fig. 4.5.

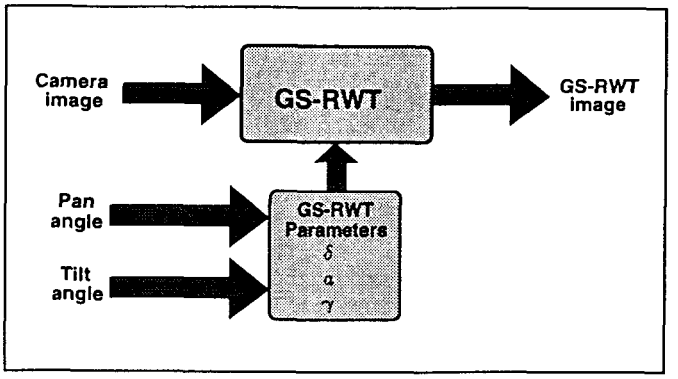


Figure 4.5: Perspective distortion compensation by the *GS-RWT*

4.6 Perspective distortion compensation

In the previous sections, image plane and object plane configurations were studied and we hypothesized that perspective distortion caused by camera pan-tilt can be compensated by a rotation in image plane, followed by a *S-RWT*, and finally a rotation in the imaginary *GS-RWT* plane. The resulting image is an canonical frontal view of the object plane. The composition of these three transforms are called the *GS-RWT*. Its matrix notation has the camera pan-tilt angles as dependent variables. In this section, the *GS-RWT* will be applied to the distorted image to prove the correctness of the hypothesis.

The *GS-RWT* is a $2D$ transform performed on the image plane. Using a homogeneous representation of the $2D$ image coordinate, the representation of an arbitrary point $[X, Y, Z_c]^T$ in the new camera coordinates after camera pan-tilt is shown in Eq. (4.1). To obtain the homogeneous representation of the projected point on image plane, it is multiplied by the projection matrix. The result is the distorted image on the image plane. The *GS-RWT* is then applied to the distorted image to project it onto the *GS-RWT* plane,

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = T_{GS-RWT}(\alpha,\beta) \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Xcos\beta + Ysin\alpha - Z_c cos\alpha sin\beta \\ Ycos\alpha + Z_c sin\alpha \\ Xsin\beta - Ysin\alpha cos\beta + Z_c \cos\alpha cos\beta \end{bmatrix} \tag{4.23}$$

$$= \begin{bmatrix} \frac{X(\sin^2\alpha\cos^2\beta+\sin^2\beta)+Z_c\cos\alpha\sin\beta\cos\beta}{\sin^2\alpha+cos^2\alpha\sin^2\beta} \\ \frac{Y(\sin^2\alpha\cos^2\beta+\sin^2\beta)-Z_c\sin\alpha\cos\alpha\cos^2\beta}{\sin^2\alpha+cos^2\alpha\sin^2\beta} \\ \frac{-Z_c}{\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}} \end{bmatrix}. \tag{4.24}$$

By Eq. (4.24), the $2D$ coordinate of the resulting image is calculated from $u_{image} = \frac{u}{w}$ and $v_{image} = \frac{v}{w}$,

$$\begin{bmatrix} u_{image} \\ v_{image} \end{bmatrix} = \begin{bmatrix} -X \cdot \frac{\sin^2\alpha\cos^2\beta+sin^2\beta}{Z_c\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}} - \frac{\cos\alpha\sin\beta\cos\beta}{\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}} \\ -Y \cdot \frac{\sin^2\alpha\cos^2\beta+\sin^2\beta}{Z_c\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}} + \frac{\sin\alpha\cos\alpha\cos^2\beta}{\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}} \end{bmatrix}. \tag{4.25}$$

From Eq. (4.25), the resulting image is an canonical frontal view of the object plane, i.e., it is only scaled and translated. The scaling factor $\frac{\sin^2\alpha\cos^2\beta+sin^2\beta}{Z_c\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}}$ is inversely proportional to the distance Z_c from camera center to the object plane and it is also affected by the camera pan-tilt angles. There is a shift in the u direction by $-\frac{\cos\alpha\sin\beta\cos\beta}{\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}}$ and a shift in the v direction by $\frac{\sin\alpha\cos\alpha\cos^2\beta}{\sqrt{\sin^2\alpha+cos^2\alpha\sin^2\beta}}$. Recalling Fig. 4.4 (b), since the origin of o_{GS-RWT} is fixed at $(1,0,0)$ with respect to the reference coordinates, the *GS-RWT* plane will be shifted relative to the object. This causes u,v shifts in the resulting image. For the special case of only camera tilt, the *GS-RWT* gives the same result as the *S-RWT*, so that Eq. (4.25) is equal to Eq. (3.18) with $\alpha = 0$.

4.7 Singularity of the GS-RWT and the Vanishing Line

As discussed in Chapter 2, a singularity of the RWT exists at $x = 0$, i.e., $u = \frac{1}{0} = \infty$ and $v = \frac{y}{0}$ [Ton95]. To eliminate the singularity, the $S\text{-}RWT$ is introduced to horizontally shift the center strip (and the rest of the image) away from $x = 0$. The $S\text{-}RWT$ has the form $u = \frac{1}{x-a}$ and $v = \frac{y}{x-a}$. For the purpose of singularity elimination, the parameter a should have 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. However, in this application, the $S\text{-}RWT$ is used for perspective distortion compensation, where the parameter a is dependent on the camera angle. Therefore, singularities will still exist at $x = a$. As illustrated in Section 4.2, the $GS\text{-}RWT$ can be decomposed into three transformations: two rotations and one $S\text{-}RWT$. Predictably, the $GS\text{-}RWT$ also may have the singularities.

Recalling the matrix notation for the $GS\text{-}RWT$ is as follows:

$$T_{GS-RWT}(\alpha,\beta) = Rot(-\gamma)\cdot T_{S-RWT}(a)\cdot Rot(\delta) \tag{4.26}$$

$$= \begin{bmatrix} \frac{\sin^2\alpha\cos\beta}{M} & \frac{\sin\alpha\cos\alpha\sin\beta\cos\beta}{M} & \frac{f\sin\beta}{\mathrm{M}} \\ \frac{\sin\alpha\sin\beta}{M} & \frac{\cos\alpha\sin^2\beta}{M} & \frac{-f\sin\alpha\cos\beta}{M} \\ \frac{\cos\alpha\sin\beta}{\sqrt{M}} & \frac{-\sin\alpha}{\sqrt{M}} & \frac{-f\cos\alpha\cos\beta}{\sqrt{M}} \end{bmatrix}$$

$$\text{where } \mathrm{M} = \sin^2\alpha + \cos^2\alpha\sin^2\beta$$

There are two situations that may cause singularities. The first situation is when $M = 0$, which makes the denominators of matrix elements zero. Since $\mathrm{M} = \sin^2\alpha +$

$\cos^2\alpha \sin^2\beta$, the pan-tilt angles both have to be zero. By assumption, for α and β zero, the camera position is such that the camera plane is parallel to the object plane; in other words, the camera image is already a canonical frontal view, and no further distortion compensation is necessary.

To explore the second situation causing singularities, we shall examine the resulting image after applying the *GS-RWT*. Let $[x, y, 1]^T$ be the homogeneous representation of a pixel on the camera image, $[u, v, w]^T$ be the homogeneous representation of a pixel on the *GS-RWT* image and $[u_{image}, v_{image}]^T$ be the pixel coordinate on the resulting *GS-RWT* image,

$$\begin{bmatrix} u_{image} \\ v_{image} \end{bmatrix} = \begin{bmatrix} \frac{u}{w} \\ \frac{v}{w} \end{bmatrix} \simeq \begin{bmatrix} u \\ v \\ w \end{bmatrix} = T_{GS-RWT}(\alpha, \beta) \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{4.27}$$

where the sign "$\simeq$" means equivalency within the homogeneous coordinate representation.

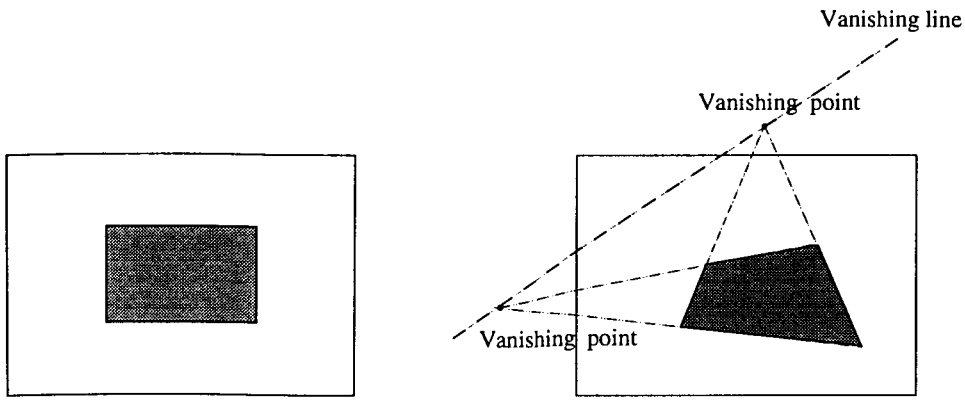
Singularity exists when $w = 0$. From Eq. (4.26) and Eq. (4.27), $w = x\frac{\cos\alpha \sin\beta}{\sqrt{M}} - y\frac{\sin\alpha}{\sqrt{M}} - f\frac{\cos\alpha\cos\beta}{\sqrt{M}}$. With $M \neq 0$, $w = 0$ determines the line

$$x = \frac{\sin\alpha}{\cos\alpha \sin\beta} y + \frac{f}{\tan\beta} \tag{4.28}$$

Therefore, given a pair of known camera pan-tilt angles α and β, the singularity points of the camera image are those points on the line given by Eq. (4.28). We will show that this line is actually the *vanishing line* for the object plane.

Vanishing line [HS93] is the set of all vanishing points for a certain plane. It is shown in Chapter 4 that if the object plane is not parallel to the camera image plane, parallel lines on the object plane will converge to a single point on the image, called

vanishing point. As illustrated in Fig. 4.6, all the vanishing points are collinear.



(a) Frontal View (b) Distorted View

Figure 4.6: The vanishing line
(a) The frontal view of a rectangular shape. (b) The distorted view of a rectangular shape. The vanishing points of parallel lines lie along the vanishing line.

A plane with surface normal (A, B, C) can be represented as $AX+BY+CZ+D = 0$ for some D. The vanishing line for such a plane is the line L_∞, where $L_\infty = Ax + By + Cf = 0$ and f is the camera focal length [HS93]. Using camera-centered coordinate system, if the camera pan-tilt angle are α and β, the surface normal for object plane is give by Eq. (4.11).

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -\cos\alpha\sin\beta \\ \sin\alpha \\ \cos\alpha\cos\beta \end{bmatrix}. \tag{4.29}$$

Substituting Eq. (4.29) into $L_\infty = 0$, the vanishing line can be represented as

$$x = \frac{\sin\alpha}{\cos\alpha\sin\beta}y + \frac{f}{\tan\beta}. \tag{4.30}$$

It is interesting to see that Eq. (4.28) and Eq. (4.30) are the same. This result

indicates that the image pixels on the vanishing line cause the singularities for the *GS-RWT*. This is due to the fact that the points on the vanishing line are at infinity. They are unreachable after the mapping. In other words, the vanishing line is the bound of the object plane on the camera image. No projection of the points on the object plane can go beyond the vanishing line.

The vanishing line is parallel to the intersection line of the camera plane and the object plane (see Eq. (4.13)). As discussed in Chapter 4, the perpendicular vector to the intersection line determines the direction of the space-variant resolution. The pixels which are on the same parallel line to intersection line have the same resolution, and resolution decreases when the parallel goes to infinity. Therefore, mathematically, the image pixels on the vanishing line have infinitesimal resolution.

Chapter 5

Applications and Experimental Results

This chapter presents some experimental results of the Generalized Shifted Reciprocal Wedge Transform for the compensation of perspective distortion and its application in mobile robot navigation. In Section 5.1, the *GS-RWT* is used to compensate the distorted images resulting from different camera pan-tilt angles. In Section 5.2, experimental results from robot navigation demonstrate the application of the *GS-RWT* to camera gaze control and object tracking and approaching.

5.1 Perspective distortion compensation by GS-RWT

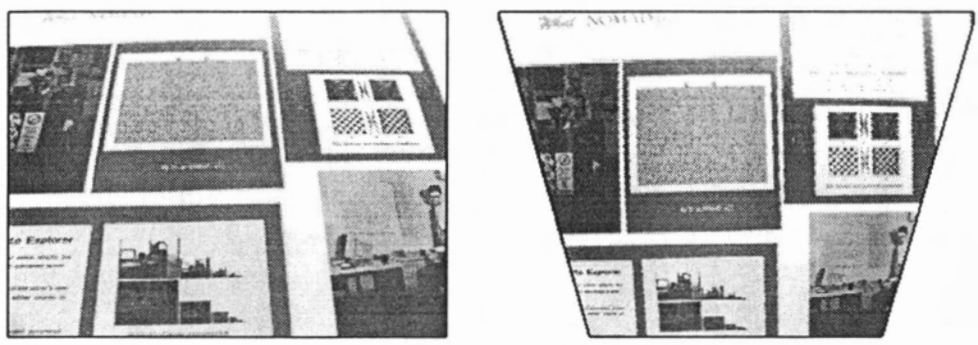
Corridor scenes taken by a CCD camera with different pan-tilt angles are shown in the left column of Fig. 5.1. The *GS-RWT* is applied to compensate perspective distortion and the resulting images are displayed in the right column. Whereas the

original images are 240×320 pixels, the compensated images have different sizes, due to the *GS-RWT*. For display convenience, the images have been scaled to have the same row number.

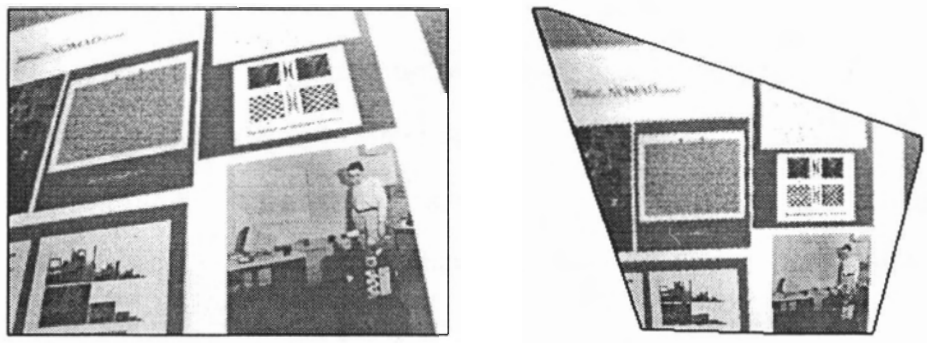
A rectangular image becomes a distorted quadrangle after the *GS-RWT*, with parallel edges transformed to converging lines. The direct implementation of forward *GS-RWT* by the Eq. (4.26) resulted in a broken, sparse mapping. However, it can be used to determine the boundary of the resulting image. The actual mapping is done by performing an inverse *GS-RWT* to map the pixel backwards.

The *GS-RWT* parameters for the set of images in Fig. 5.1 are shown in Table 5.1. The camera-focal-length-related constant was calibrated to be 280 in terms of image pixels. The direction of space-variant resolution was determined by the rotation angle δ and the second rotation angle γ was used to restore the canonical frontal view. Angles δ and γ were calculated from the required camera pan-tilt. For example, in Fig. 5.1(a), the camera is tilted, and an *S-RWT* is sufficient to compensate for distortion, therefore, $\delta = 0°$. Moreover, the image after applying the *S-RWT* is already a canonical frontal view, so that $\gamma = 0°$. Fig. 5.1(d) is taken from a camera position with pan angle $\alpha = 30°$ and tilt angle $\beta = 30°$. Consequently, $\delta = 49°$ and $\gamma = -41°$ were needed to perform the proper *GS-RWT*.

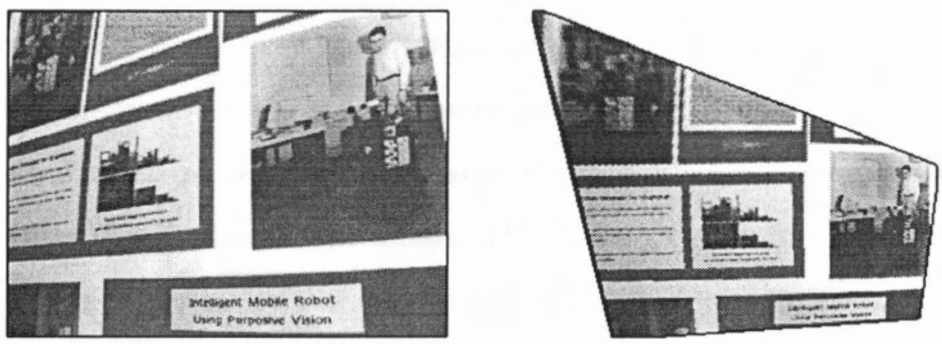
The vanishing point of the parallel lines which are perpendicular to the intersection line is called *main vanishing point.* For the special case of only camera tilt, the main vanishing point is the converging point of the vertical lines. The shift parameter a is the displacement of the main vanishing point. The positions of the main vanishing points in Table 5.1 are expressed in terms of image pixels with the origin at the center of the camera image. Since the images in Fig. 5.1 were obtained with non-negative



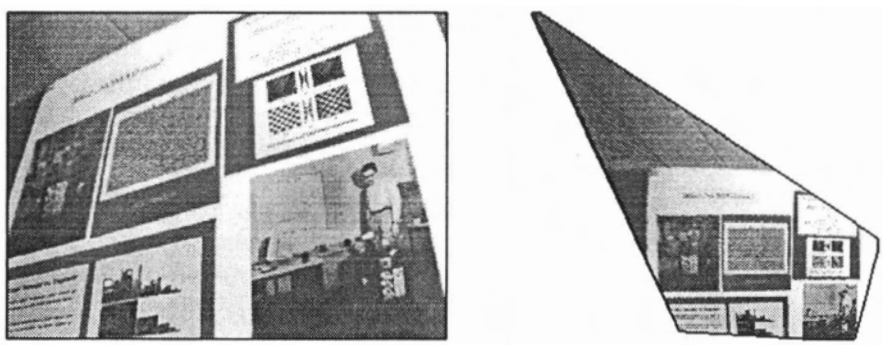
(a) pan angle = 0° and tilt angle = 30°



(b) pan angle = 15° and tilt angle = 30°



(c) pan angle = 30° and tilt angle = 15°



(d) pan angle = 30° and tilt angle = 30°

Figure 5.1: Perspective distortion compensation for different camera pan-tilt angles

Table 5.1: Parameters of the *GS-RWT*

Items	Image a	Image b	Image c	Image d
Pan angle	0°	15°	30°	30°
Tilt angle	30°	30°	15°	30°
δ	0°	28°	66°	49°
Main vanishing point	$(0, 485)$	$(-201, 376)$	$(-390, 174)$	$(-240, 207)$
a	485	427	427	317
γ	0°	$-24°$	$-61°$	$-41°$

pan-tilt angles, the main vanishing points are located in the second quadrant. The magnitude of a indicates the rate of space-variant resolution. As discussed in Section 4.7, the vanishing line passes through the main vanishing point. The points on the object plane become compressed when they are close to the vanishing line. The *GS-RWT* reverses this process by decompressing the compressed portion. For example, Fig. 5.1(d) has the smallest a among the four images. We can observe that the top-left part of the image has been greatly decompressed after use of the *GS-RWT*.

The problem occurs when the camera is pan-tilted such that the camera image plane and object plane have a very large intersection angle. In these cases, the vanishing line is very close to the image. This causes over-decompression of the image portion which is close to the vanishing line. As illustrated in Fig. 5.2, the camera image is taken with the pan and tilt angles 45° and 30°, respectively. The main vanishing point is at $(-193, 96)$ and the shift parameter is $a = 217$. The upper-left corner is over-decompressed so that it occupies most of the resulting image. Proportionally, it occupies most of the area of the *GS-RWT* image and the useful, central part of the image is squeezed to the bottom-right corner.

The extreme occurs when the vanishing line intersects the image. Therefore, the

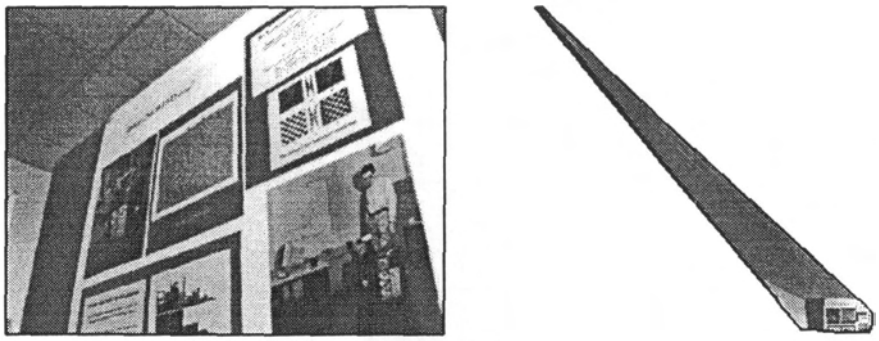


Figure 5.2: Over-decompressed image

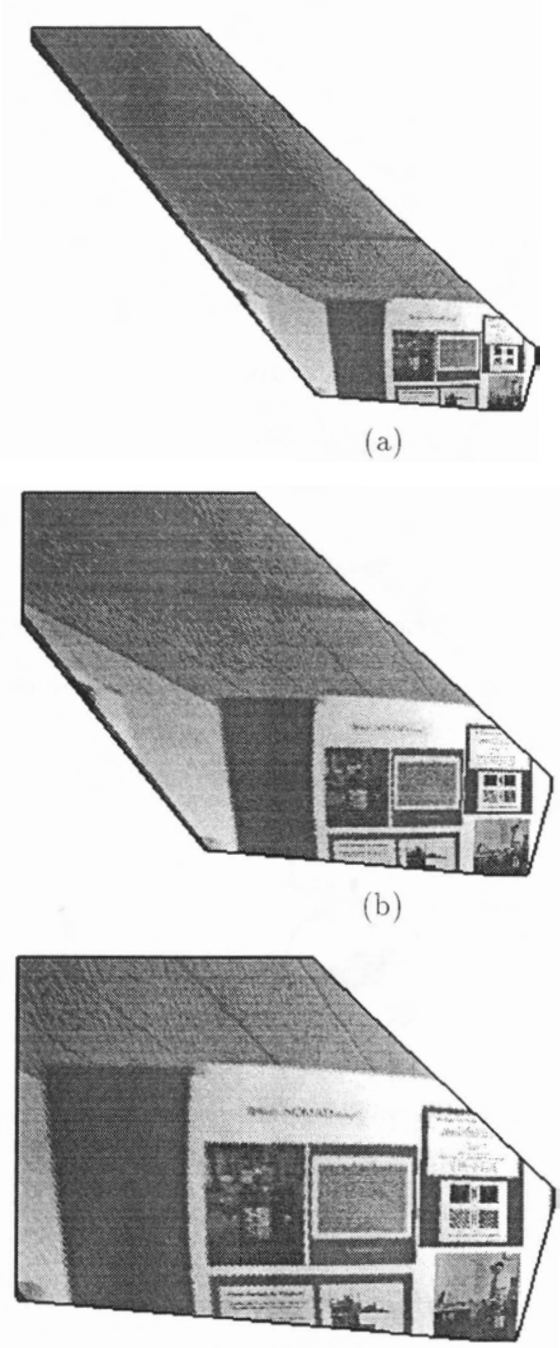
image pixels on the vanishing line are singularity points which will cause computational difficulties when the *GS-RWT* is applied to these pixels.

A remedy for this difficulty is the application of a bounding box on the *GS-RWT* image. The over-magnified parts of the images are those close to the vanishing line. In other words, they are pixels on the object plane which are very far away from the camera and compressed into small regions in the distorted image. The *GS-RWT* compensates the distortion by decompressing these regions. However, these regions are less interesting to the viewer since they are too far away and the resolution is too low to be processed. Fig. 5.3 illustrates the resulting images with different size bounds where the size of the bounds box decreases from top to bottom.

5.2 The GS-RWT for mobile robot navigation

5.2.1 System structure

The system is based on a wheeled mobile robot (Nomad 200, Nomadic Technologies Inc.), a computer controlled pan-tilt mechanism (Model PTU, Directed Perception) and a CCD camera (TI-23A, NEC) mounted on the top of the pan-tilt unit. The



(a)

(b)

(c)

Figure 5.3: The *GS-RWT* images with a bounding box

Nomad 200 has an ultrasonic module consisting of a ring of 16 sensors that yields the distance to the objects detected by a sonar.

The pan-tilt unit has two degrees of freedom. It can rotate about the horizontal and vertical axis. Motion is controlled by an on-board 486 PC through an RS-232 serial port. A 2D image of the environment is captured by image grabber hardware, through which the visual image is stored as a two dimensional gray-level array. The resolution of the image is determined by the particular task. A low resolution image is sufficient for navigation and door detection. A high resolution image is necessary for detailed object recognition.

Ultrasonic sensor data and visual image inputs offer two sources of perception for the system. Integration of these two sensory inputs provides the knowledge of the environment that is required for the navigation task (see. Fig. 5.4).

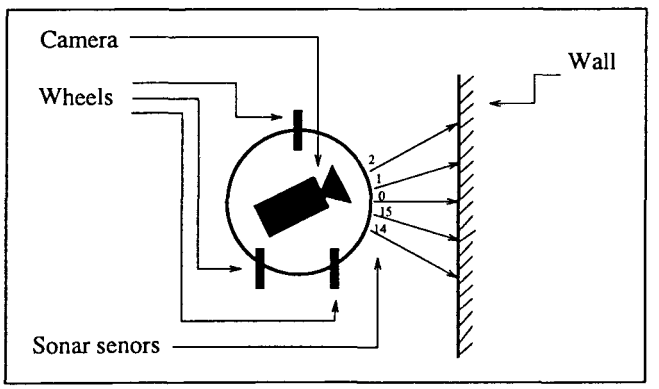


Figure 5.4: Mobile robot in navigation

The turret of the robot is locked during the navigation. The forward direction is determined by the rotation of the wheels and it remains parallel to one side of the wall. The direction of Sonar sensor 0 (see Fig. 5.4) is kept perpendicular to the wall. By measuring the distance input from sensor 0, the robot maintains a

certain distance from the wall. To move parallel to the wall, four other ultrasonic sensors are employed. Since the mobile robot is cylindrical, the ultrasonic beam from each sensor can be considered to emit from the center of the body. To maintain the forward direction of the robot parallel to the wall, distance readings from sensors 1 and 15 should be identical and so are the reading from sensors 2 and 14. Based on the difference of readings from these two sensors, the robot posture is adjusted. The wheels are maintained parallel to the wall in order to keep the same posture.

A corridor is often bounded by two vertical walls. Navigation in this environment is based on the successful recognition of the landmarks by the active vision. The above method offers a good estimation of robot posture with respect to the wall. The recognition of the landmarks is refined by spatial and angular measurements of the environment as described below.

5.2.2 Camera gaze control

A door detection task demonstrates landmark recognition with gaze control. A door is characterized by left, right and top door boundaries and the door knob. Other features such as a pneumatic damper and peek window may also be utilized. Their relative locations are always the same. For example, left and right boundaries have a certain distance between them, and the knob is close to one of the boundary. Door detection, therefore, can be achieved by recognition of these sub-objects.

Gaze change is context-driven. It is based on the results of sub-object recognition. Fig. 5.5 illustrates a sequence of the gaze changes and Fig. 5.6 displays the camera image taken from six fixation points. In Fig. 5.6(a), a door-boundary-like vertical line is seen by the robot which is traveling down the corridor from the left to right.

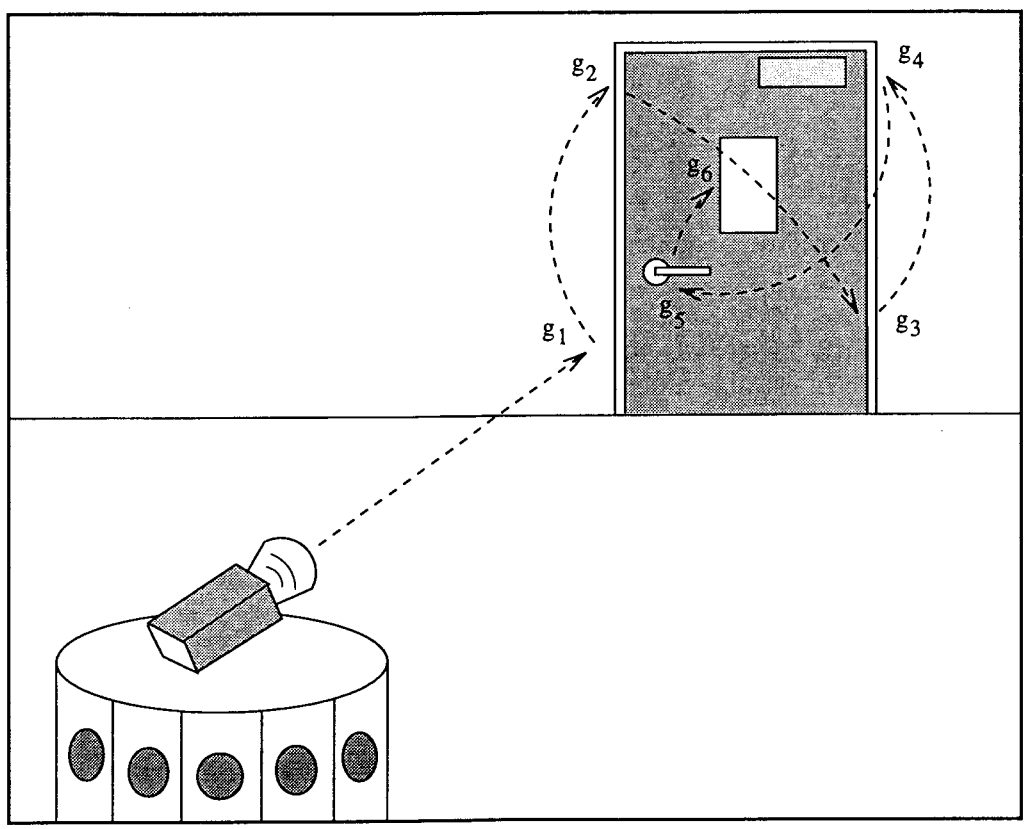


Figure 5.5: Gaze change for door detection

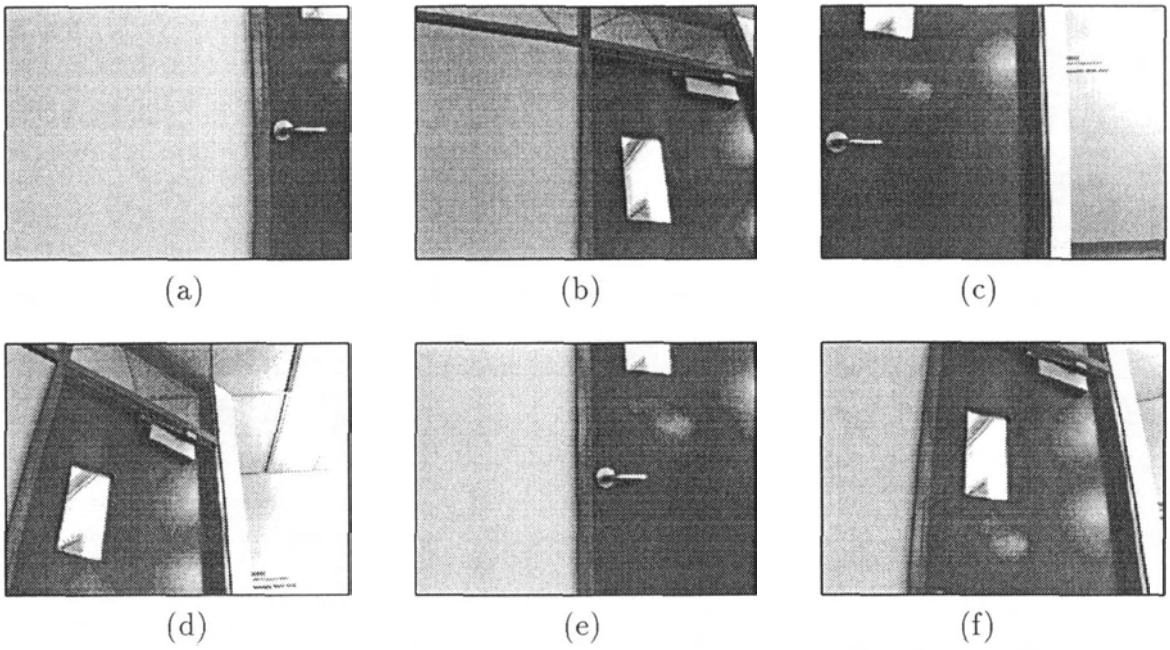


Figure 5.6: Images taken from the six fixation points for door detection

In Fig. 5.6(b), camera is tilted up to detect the upper part of the vertical line and ensure the length of the line. After the first two steps, the long vertical line can be assumed to be the left boundary of the door. To gather more information, camera is pan-tilted to check the right and top door boundary in a similar way in Fig. 5.6 (c) and (d). Through the previous steps, adequate evidences are obtained to prove the existence of the door. Further checking is performed in Fig. 5.6(e) and (f) to detect the door knob and the peek window.

The original camera images suffer from perspective distortion in the following two ways. First, the shape of the object varies from the model under different camera pan-tilt angles which makes a simple, direct model matching infeasible. Second, the relative positions of the sub-objects are not preserved in the distorted image so that it affects the context-driven gaze change for the next probe point.

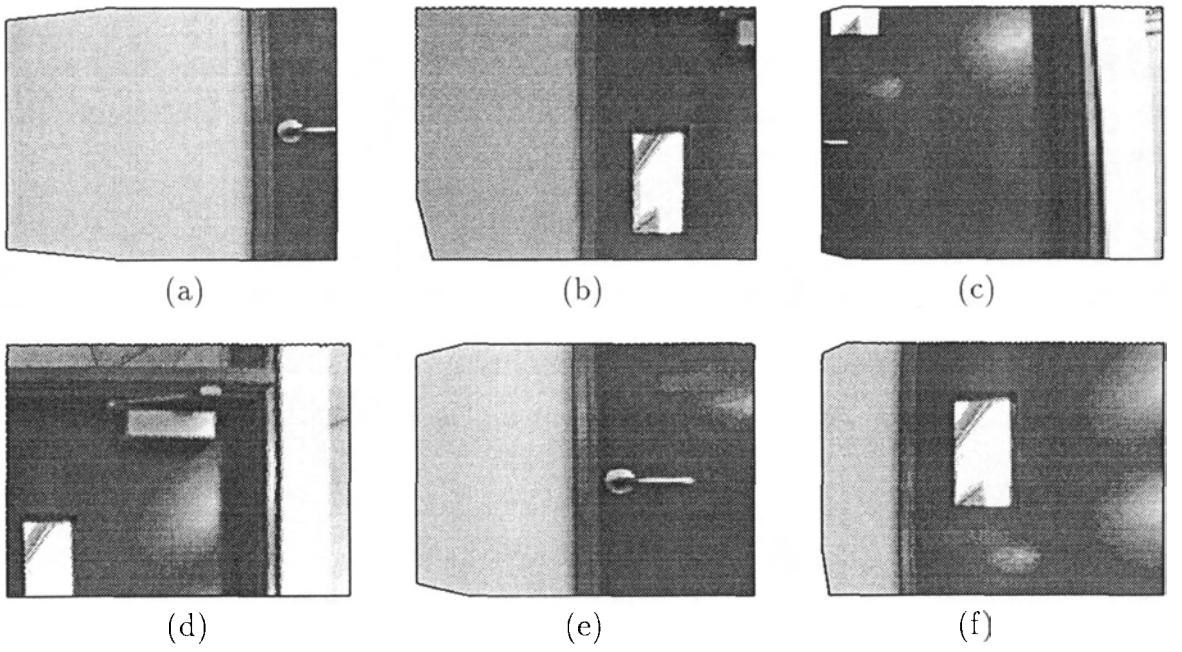


Figure 5.7: The *GS-RWT* images for the six fixation points

The *GS-RWT* images have a canonical frontal view by the elimination of perspective distortion. The corresponding *GS-RWT* images for every probe point are shown in Fig. 5.7. Bounds are applied to retain only the central part of the image as shown. The selected region is the rectangular area seen by the camera when it is aimed directly at the wall. The center of the bounding box is the point where the camera optical axis hits the wall plane. It can be calculated by applying the *GS-RWT* to the center pixel of the image. The *GS-RWT* images are 120×160 pixels.

The Hough Transform line detection algorithm [DH72] is used to find the door boundaries. The Hough Transform maps straight lines in an image to a parameter domain. The cell of the parameter domain is indexed by its normal θ and its algebraic distance ρ from the origin. The equation of a line corresponding to this geometry is:

$$x \cdot \cos\theta + y \cdot \sin\theta = \rho.$$

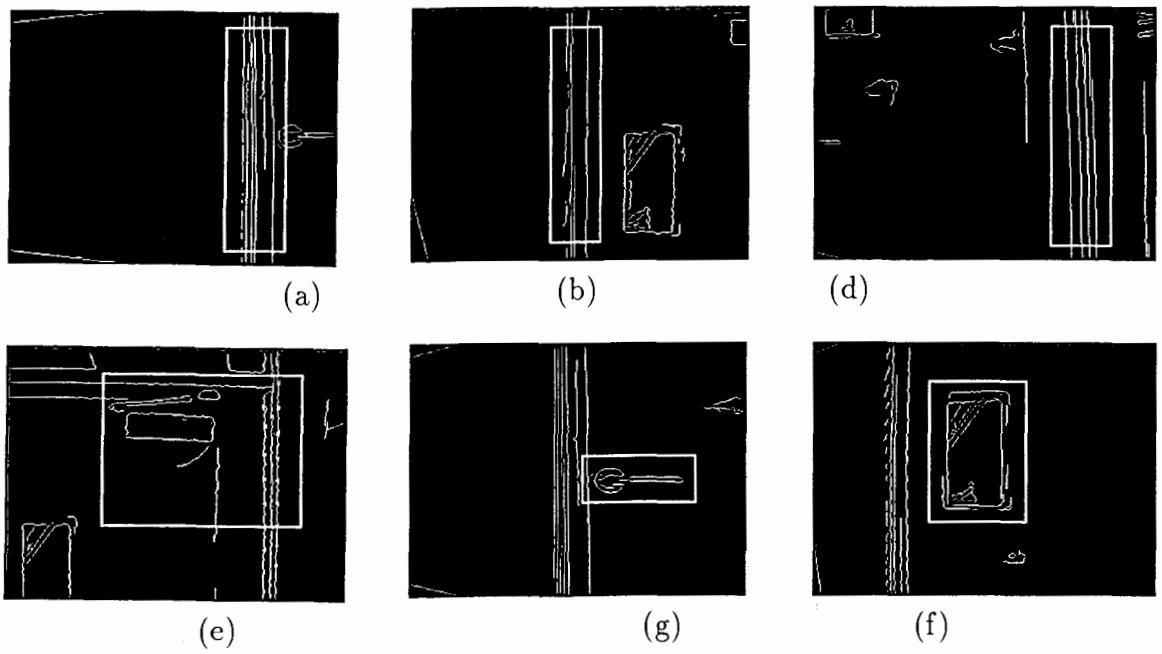


Figure 5.8: Door detection by Hough Transform

Image pixels vote in the parameter domain. The cells of the parameter domain with adequate number of votes are considered as lines. A threshold of 50 is used as a criteria of line.

Gradient magnitudes and orientations of the images pixels are calculated by applying a 3×3 sobel mask. Gradient magnitudes are thresholded to indicate the edge pixels. The thresholding value used is 80 in terms of pixel grey level. The gradient orientation is equal to the line normal θ. Its value is within 0°–360° and sampled for every 2°. The door boundaries detected by the Hough Transform are shown in Fig. 5.8 (a), (b), (c) and (d).

Since the *GS-RWT* has the property of shape-preserving, object detection is performed using the General Hough Transform [Bal81] with a uniform model on the *GS-RWT* image. Similar to the Hough Transform, the pixel on the object model is

represented by the distance and angle to a reference point of the object and indexed by the gradient orientation. The models of the door knob and peek window are generated off-line and used throughout the navigation task. Modeling is performed within a bounding box which contains the object and the center of the bounding box is used as the reference point of the object. All the edge pixels inside the bounding box are considered as part of the model. The bounding box for the door knob is of the size 40×100 and the one for the peek window is 120×90. The locations of the closest matching for these two objects are shown in Fig. 5.8 (e) and (f).

The vector distances between the feature objects are hard-coded as part of the model information. Since the *GS-RWT* offers a canonical frontal view of the scene, the vector distances can be directly used to guide the gaze control. The resulting *GS-RWT* image only differs from the world scene by a scale factor which can be calculated from Eq. (4.25). The scale factor is a constant for a certain camera fixation. Therefore, a new fixation position can be estimated by a simple vector calculation on the *GS-RWT* image, which is subsequently converted to camera pan and tilt values. For example, the camera pan-tilt values for Fig. 5.7(b) are $-15°$ and $30°$, respectively; those for Fig. 5.7(d) are $-35°$ and $26°$, respectively.

5.2.3 Object tracking and approaching

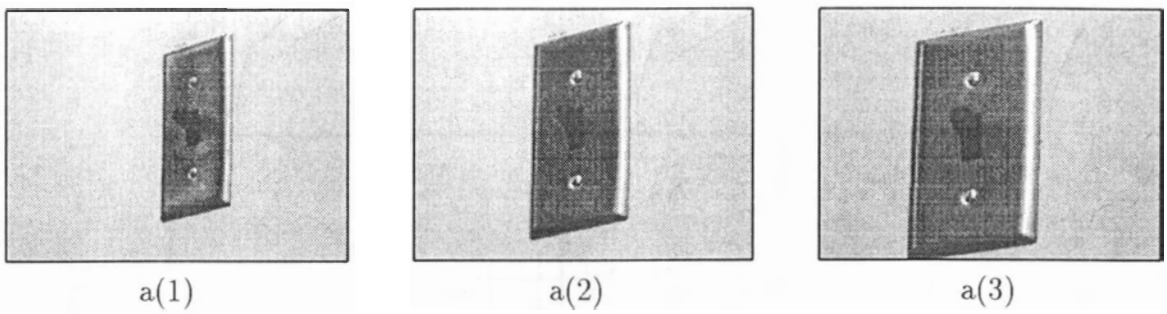
Object tracking consists of controlling the robot movement and camera pan-tilt to maintain good views of the object of attention. In our navigation task, all landmarks are static objects. However, due to the robot movement, camera pan-tilt angles must be adjusted to keep track of the landmark of interest.

A high-resolution image of the object is often preferable for a detailed examination

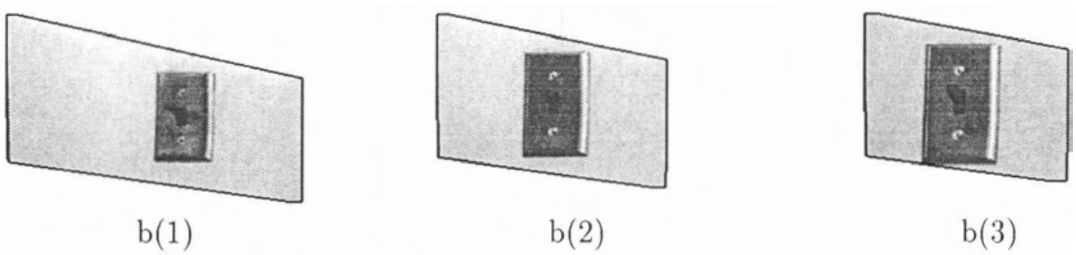
of the scene. In our task, a uniform resolution camera was used. A high resolution image is achieved by movement of the robot itself and object tracking during the approaching process. For example, suppose there is an object of interest on one side of the wall in the distance. To obtain a better view, the robot moves parallel to the wall, approaches the object, and keeps the object in the camera field of view.

Alerting and object tracking is based on object matching. The *GS-RWT* is used to compensate the perspective distortion. By a similar argument as in the previous section, the location of the object is simplified by the shape-preserving feature of the *GS-RWT*. The required camera pan-tilt angles are calculated from the relative location of the object to the robot and the robot navigation velocity. Suppose the robot is directed to find a switch panel on the wall (See Fig. 5.9). Three snapshots are taken during the approaching. The robot moves down the corridor and parallel to the wall with a distance of 25 inches. The switch panel is on side of the wall and approxiately 33 inches away. The first snapshot is taken with the camera angle 53° and 13°. The robot moves forward for approximately 12 inches. In order to track the object, the camera pan-tilt angles are calucualted as 40° and 15°. The third snapshot is taken after another forward movement of 6 inches and with the camera pan-tilt angle 29° and 18°. The resolution of the switch panel improves as the robot moves closer to the object of interest.

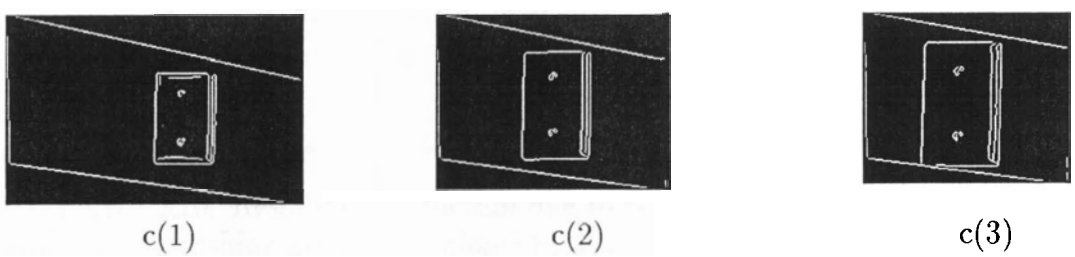
The higher resolution of the object image gained by the approaching movement is based on the following two facts. First, the approaching of the robot shortens the absolute distance between the object and the camera which results in a larger appearance of the object. Second, and more subtly, approaching of the robot reduces the viewing angle which also improves the image resolution.



a. The original camera images



b. The *GS-RWT* images with the same resolution



c. Edge maps of the *GS-RWT* images

Figure 5.9: Object tracking and approaching

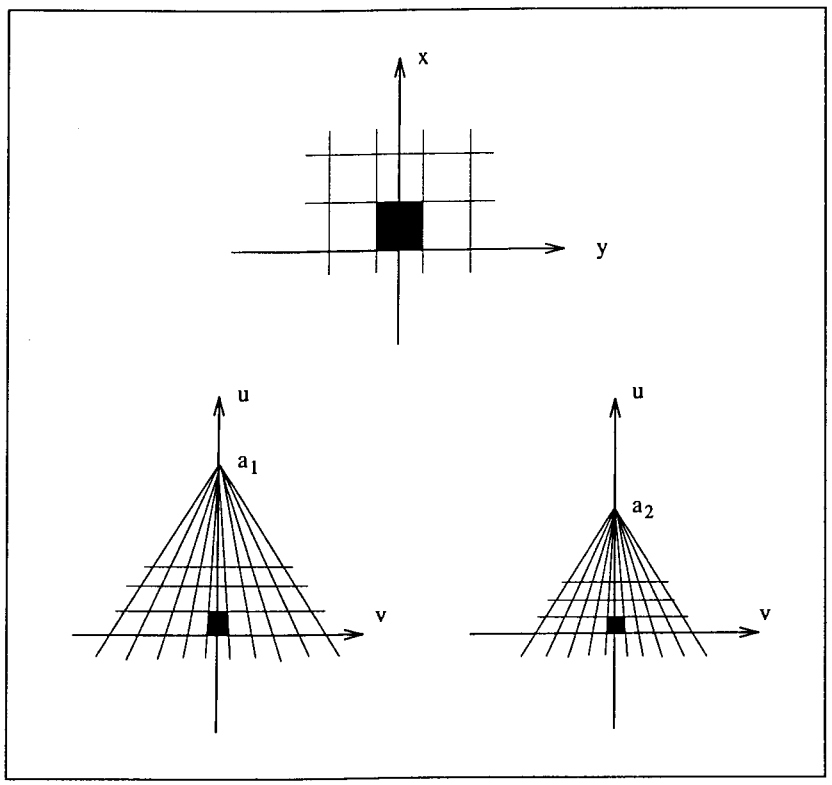


Figure 5.10: Resolution reduction due to the position of the vanishing point
The main vanishing point a_2 is closer to origin than a_1. Therefore, the image pixels with the same coordinate in x-y domain become more compressed in the distorted image.

Table 5.2: Resolution reduced due to camera pan-tilt angle

Items	Image 1	Image 2	Image 3
Pan angle	53°	40°	29°
Tilt angle	13°	15°	18°
Interception angle θ	54.10°	42.27°	33.71°
Jacobian determinant	0.59	0.74	0.83

By maintaining the object in the camera field of view, the approaching movement reduces the interception angle θ of the image plane and the wall plane. The interception angle is also the viewing angle of the camera. If it increases, the camera view is more oblique. The angle θ is discussed in Chapter 4. It is the composite angle of camera pan and tilt. The main vanishing point occurs at $a = \frac{f}{\tan\theta}$ along the direction of space-variant resolution. Therefore, the greater θ, the closer a is to the origin, and vice versa. Since the resolution varies depending upon θ only, without loss of generality, Fig. 5.10 illustrates the situation when a is on u axis. If a is closer to the origin, the resolution of the image pixel becomes more reduced. A similar situation occurs when the direction of resolution reduction is away from u axis.

In Table 5.2, the Jacobian determinants [Ton95] are calculated to indicate the area of the pixel reduced at the origin. The Jacobian determinants are compared to the area of the correspondent pixel in the frontal image. It shows that the Jacobian determinant increases when the θ angle decreases. In other words, the resolution of the image improved when the viewing angle is smaller. The *GS-RWT* images shown in Fig. 5.9 are computed so that the resolution remains consistent with that in original camera image. The resolution gradually improves, when the robot approaches closer. This imitates the human capability of changing from coarse to fine as an object is

approached.

Chapter 6

Conclusions and Discussion

In this thesis, we have developed an image representation model called the Generalized Shifted Reciprocal-Wedge Transform ($GS\text{-}RWT$). This technique is an extension of the Reciprocal-Wedge Transform (RWT) model due to Tong. The space-variant resolution and shape-preserving properties of $GS\text{-}RWT$ are addressed and used in the compensation of the perspective distorted images. This technique is applied in mobile robot navigation and demonstrated to simplify the robot gaze control, object tracking and approaching. This chapter summarizes the contributions and suggests some extensions for future research.

6.1 Contributions

1. The Generalized Shifted Reciprocal-Wedge Transform ($GS\text{-}RWT$) is developed as an image model for space-variant representation.

   The $GS\text{-}RWT$ can be represented in the matrix notation as a composite transform of one $S\text{-}RWT$ and two rotations. The anisotropic variable resolution is

derived from the *S-RWT* in which the shift constant a is used to adjust the compression rate. The two rotations enable the *GS-RWT* to have the principle direction of variable resolution away from x-axis. The *GS-RWT* inherits the linearity-preserving property of the *RWT*, thus the generalization of the *S-RWT*. The projective model of the *GS-RWT* is developed to geometrically relate the *GS-RWT* to the V-plane projection. The notion of a vanishing line is introduced to describe the singularity of the *GS-RWT* and a method using a bounding box is applied to remedy the singularity.

2. The *GS-RWT* is shown to be suitable for compensating the perspective distorted images.

   The compensation of the perspective distortion is motivated by the fact that the image plane of the pan-tilted camera forms a V-plane projection with the object plane. It is shown that the distortion caused by camera tilt can be compensated by the use of the *S-RWT*. The shift constant a of the *S-RWT* is determined by the camera tilt angle. It is also shown that the *GS-RWT* can compensate the distortion caused by camera pan and tilt. It is shown that the parameters of the *GS-RWT* can be derived from the camera pan and tilt angles. The resulting image is a canonical frontal view of the distorted image and the original shape of the object is preserved.

3. The *GS-RWT* is demonstrated in a mobile robot task for gaze control and object tracking and approaching.

   Mobile robot navigation relies heavily on landmark recognition. In a corridor environment, notable objects and signs are often on surrounding vertical surfaces

such as walls and doors. With the shape-preserving property, the *GS-RWT* is applied to camera gaze control and object tracking and approaching tasks of robot navigation. Gaze control of the camera is simplified because the pan-tilt angles can be directly measured from the canonical frontal view. Object tracking and approaching is facilitated since object recognition is performed on the recovered uniform resolution images and image resolution is improved during the approaching phase.

6.2 Future research

The *GS-RWT* preserves linear features in the image, and its anisotropic variable resolution is suitable for tasks that suffer from perspective distortion. Moreover, the *GS-RWT* can be generally viewed as an alternative model to space-variant sensing such as the log-polar transform. Some suggestions are made in the following as extensions of this work or future directions to related work.

1. Automatic generation of the *GS-RWT* parameters for mobile robot navigation.

   In our robot navigation application, the parameters of the *GS-RWT* are determined from the camera pan-tilt angles with respect to one side of the wall. The robot self-positioning and distance measurement rely on the range readings from the ultrasonic sensors. This approach is suitable for well-defined environments, such as a corridor that consists of two long, smooth vertical walls. In such cases, ultrasonic sensors can provide a reasonably accurate measurement. However, this approach may not be suitable for more cluttered environments. In a office, for example, there are many obstacles, such as tables and chairs.

Ultrasonic sensors may only be used for obstacle avoidance. They are not useful for wall orientation detection. Consequently, the camera pan-tilt angles cannot be estimated.

However, a lot of visual cues can be learned by active observation of the surrounding environment. For example, the location and orientation of the vanishing line of an object plane determines the true 3D orientation of the wall with respect to the camera coordinates. Therefore, the *GS-RWT* parameters can be automatically generated from visual feedback, rather than the knowledge of the camera pan-tilt angle. This capability significantly extends the usage of the *GS-RWT* model.

2. Active vision with foveal sensing.

The *GS-RWT* has the property of anisotropic variable resolution. The principle axis of space-variant resolution is determined by the rotation angle δ. By applying the *GS-RWT*, the image pixels close to the vanishing line are magnified and have a high resolution, and vice versa. Therefore, the region close to the vanishing line can be considered as the fovea of the space-variant vision system and the rest of the image is the periphery.

Two key issues in active vision are as follows: (1) Space-variant sensors must be fixated on task-specific regions of interest in order to utilize the higher-resolution parts of the sensor and, (2) Image processing and pattern recognition methods appropriate to space-variant sensors must be developed. In our applications,

camera gaze control and object tracking and approaching benefit from the shape-preserving property of the *GS-RWT*. The pattern-recognition problem in space-variant sensing is solved. However, foveal sensing is not fully implemented. Further research in this direction can be done to combine these two technologies.

Bibliography

[Alo90] Y. Aloimonos. Purposive and qualitative active vision. In *Proc. Image Understanding Workshop*, pages 816–828, 1990.

[Alo93] Y. Aloimonos. Introduction: Active vision revisited. *Active Perception*, 13:4–16, 93.

[AWB88] Y. Aloimonos, I. Weiss, and A. Bandopapadhay. Active vision. *Int'l. J. Comp. Vision*, 7:333–356, 1988.

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

[Bal81] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111–122, 1981.

[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.

[BPK94] M. Bishay, R.A. Peters, and K. Kawamura. Object detection in indoor scenes using log-polar mapping. *Proc. IEEE*, pages 775–780, 1994.

[BR92] J.R. Beverige and E.M. Riseman. Hybrid weak-perspective and full-perspective matching. In *Proc. IEEE Int. Computer Vision and Pattern Recgnition*, pages 432–438, 1992.

[BWR90] J.R. Beverige, R. Weiss, and E.M. Riseman. Combinatorial optimization applied to variable scale 2d model matching. In *Proc. IEEE Int. Conf. on Pattern Recognition*, pages 18–23, 1990.

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

[CMST87] C. Connolly, J. Mundy, J. Stensrom, and D. Thompson. Matching from 3-d range models into 2-d intensity scenes. In *Proc. IEEE Int'l. Conf. Robotics and Automation*, pages 65–72, 1987.

[DH72] R.O. Duda and P.E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. Assoc. Comput. Mach.*, 15:11–15, 1972.

[Gri90] W.E.L. Grimson. *Object recognition by computer: The role of geometric constraints.* MIT Press, 1990.

[Ham92] T. Hamada. Active vision. In *Proc. Int'l. Neuroethology Congress*, 1992.

[HS93] R.M. Haralick and L.G. Shapiro. *Computer and robot vision*, volume 2. Addison-Wesley, 1993.

[Kan88] K. Kanatani. Constraints and length and angle. *Computer Vision, Graphics, and Image Processing*, 4:28–42, 1988.

[Kan93] K. Kanatani. Mobile robot navigation using neural networks and nonmetrical enviroment models. *IEEE Control Systems Magazine*, 13:30–39, 1993.

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

[Kro93] E. Krotkov. Active vision for reliable ranging: Cooperative focus, stereo, and vergence. *International Journal of Computer Vision*, 11(2):187–203, 1993.

[LDW92] J.J. Leonard and H.F. Durrant-Whyte. *Directed sonar sensing for mobile robot navigation.* Kluwer Academic Puslishers, 1992.

[Low87] D.G. Lowe. Three-dimensional object recognition from single two dimensional images. *Artificial Intelligence*, 31:355–395, 1987.

[LTR95] Z.N. Li, F. Tong, and X.O. Ren. Applying reciprocal-wedge transform to ego motion. In *Proc. IASTED International Conference on Robotics and Manufacturing*, pages 256–259, 1995.

[MK86] M. Meng and A.C. Kak. Camera rotation invariance of image characteristics. *Computer Vision, Graphics, and Image Processing*, pages 266–271, 1986.

[MK93] M. Meng and A.C. Kak. Mobile robot navigation using neural networks and nonmetrical environment models. *IEEE Control Systems Magazine*, 13:30–39, 1993.

[RGL95] X.O. Ren, W.A. Gruver, and Z.N. Li. An intlligent service robot using purposive vision. In *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, volume 4, pages 3684–3689, 1995.

[Saw92] H.S. Sawhney. PhD thesis, University of Massachusetts, 1992.

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

[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.

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

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

[TL94] F. Tong and Z.N. Li. Reciprocal-wedge transform in motion stereo. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 1060–1065, San Diego, 1994.

[TL95a] F. Tong and Z.N. Li. A camera model for reciprocal-wedge transform. *Image and Vision Computing*, (accepted for publication), 1995.

[TL95b] F. Tong and Z.N. Li. Reciprocal-wedge transform for space-variant sensing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):500–511, 1995.

[Ton95] F. Tong. *Reciprocal-wedge transform: a space-variant image representation.* PhD thesis, Simon Fraser University, 1995.

[Tso92] J.K. Tsotsos. On the relative complexity of active vs. passive visual search. *International Journal Computer Vision*, 17(2):127–141, 1992.

[WB94] L.E. Wixson and D.H. Ballard. Using intermediate objects to improve the efficiency of visual search. *International Journal of Computer Vision*, 12(2/3):209–230, 1994.

[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.

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