

Compensating Perspective Distortion Caused by Camera Pan-tilt Movements

Ze-Nian Li * and Simon X.O. Ren
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada V5A 1S6

Abstract

Modern mobile robots often use visual feedback in navigation. When a robot actively uses its camera's pan-tilt movements to obtain desirable views, the shape of the object in the view is usually perspective distorted. This paper presents an image transformation — the Generalized Shifted Reciprocal-Wedge Transform (*GS-RWT*) for compensating the perspective distortions. The *GS-RWT* is an extension to the Reciprocal-Wedge Transform (*RWT*), a new space-variant image representation. First, it is shown that the distortion caused by the camera tilt can be compensated using the Shifted *RWT* (*S-RWT*). Second, the Generalized *S-RWT* (*GS-RWT*) is employed to compensate the distortion caused by both the camera pan and tilt. The resulting image is a canonical frontal view. 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.

1 Introduction

This paper deals with vision for robot navigation in simple indoor environments such as a corridor or an office. It is assumed that a camera and a pan-tilt unit is used. In general, perspective distortion is caused by the fact that the camera is not very far away from the object and hence it has a perspective view. In model-based approaches to 3D vision, it is common to describe the 3D object from a few favorable views, which can be called *canonical views*. Some of the possible canonical views will be from the top, side, and front. When a mobile robot is placed in a corridor environment, notable objects and signs are often on the surrounding surfaces such as wall, door, floor, and ceiling. When the camera on the robot is pointing directly at one of these sur-

faces (i.e., the camera's optical axis is perpendicular to the object plane), it is said to have a *frontal view* of that surface. If the frontal view happens to be one of the views that the object is modeled in, then it is a *canonical frontal view*. When the camera mounted on the pan-tilt platform on the robot is actively changing its pose, it is clear that it will not always have the canonical frontal view of the objects. The perspective distorted object in the image will not match with the models in the canonical views. In this paper, we will develop techniques for compensating the perspective distortions caused by the camera pan and tilt movements.

Kanatani [1] introduces projective geometry from a computational point of view. It is argued that vanishing points of parallel lines and vanishing lines of planar surfaces are essential in the analysis of perspective projected images. It is shown that some 2D transformations of image coordinates can convert a distorted image to its canonical frontal view, which simulates the camera rotation.

Tong and Li [2] present the *Reciprocal-Wedge Transform* (*RWT*) for space-variant imaging. The *RWT* is capable of compensating the perspective distortion while reducing the unnecessary details from its variable-resolution image, which renders it suitable for various robotics applications. Its distinctive advantage over the log-polar transform is that it preserves linear features. Recently, it is also reported [3] that a hardware *RWT* camera can be built based on the transform's simple and powerful projective model.

This paper presents 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 pre-

*This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the grants OGP36726 and EQP156328.

sented to compensate the distortion caused by both the camera pan and tilt. 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.

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.

2 Review of RWT

2.1 Definition of the RWT

The *RWT* [2] maps a rectangular image into a wedge shaped image. It is defined as a mapping of the image pixels from the x - y domain to a u - v domain such that

$$u = 1/x, \quad v = y/x. \quad (1)$$

The *RWT* defined in Eq. (1) can be represented in a matrix form by using the homogeneous coordinates, it is a cross-diagonal matrix of 1's, and the inverse of the *RWT* matrix \mathbf{T} is \mathbf{T} itself.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{T}^{-1}, \quad (2)$$

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

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

2.2 Shifted Reciprocal-Wedge Transform (S-RWT)

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

$$u = 1/(x+a), \quad v = y/(x+a). \quad (3)$$

Both the forward and backward transformations for the *S-RWT* have the same cross-diagonal matrix Eq. (2) 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

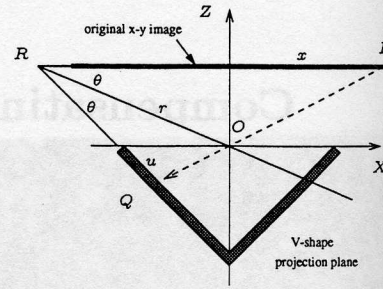


Figure 1: Geometry of the V-projection.

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.

2.3 S-RWT and V-plane Projection

The *S-RWT* can be implemented with a V-plane projection in which the original x - y image is projected onto a V-shape projection plane to obtain the corresponding RWT image. Fig. 1 depicts the V-plane projection, the two projection planes are joined to form the V shape. 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 shown that the V-plane projection implements the space-variant resolution of the *S-RWT* of Eq. (3). As illustrated in Fig. 1, the mapping of a point P on the original image plane to the projection point Q on the projection is the following:

$$u = \frac{f^2}{x+a}, \quad (4)$$

$$v = \frac{fy}{x+a}, \quad (5)$$

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

It is clear that the difference between the *S-RWT* defined in Eq. (3) and the Eqs. (4) and (5) from the V-plane projection is merely a constant factor of f or f^2 . The shift parameter a depends solely upon the V-plane configuration.

3 S-RWT for Compensating Distortion by Camera Tilt

3.1 Vanishing Point and Camera Pan-Tilt Model

In a pin-hole camera model, a 3D point (X, Y, Z) can be projected onto a point (x, y) on the image:

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (6)$$

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. 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, i.e., the *perspective distortion*.

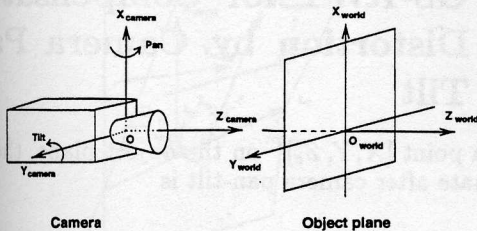


Figure 2: Camera pan-tilt model

The camera pan-tilt model is illustrated in Fig. 2. Note, in this setting camera X -axis is vertical and camera Y -axis is horizontal¹.

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 ,

$$T_{tilt}(\beta) = Rot_Y(\beta).$$

Apparently, pan is rotation about X when the camera tilt angle is zero. If the tilt angle β is not zero, then

$$T_{pan}(\alpha) = Rot_Y(\beta)Rot_X(\alpha)Rot_Y(-\beta).$$

The end position of the camera, or so-called camera pose, after the pan-tilt motions can be described by a pair of pan-tilt position angles α and β with respect to the original position. From the above, it can be readily shown that the order of the pan and tilt movements is not important as one would expect. Namely, their composite effect is always $Rot_Y(\beta) \cdot Rot_X(\alpha)$.

3.2 Perspective Distortion Compensation by the S-RWT

When the camera Z -axis is perpendicular to the object plane and the pan and tilt angles are zero, it has a canonical frontal view. Any camera pan-tilt will cause the loss of the canonical frontal view. The process of converting a perspective distorted view to a canonical frontal view is called *perspective distortion compensation*. In this section, a method based on the *S-RWT* is developed to compensate the distortion caused by the camera tilt.

¹The X -axis has been used as the principle direction of the space-variant resolution [2]. Hence, in certain applications, e.g., in this paper, it would refer to the vertical axis.

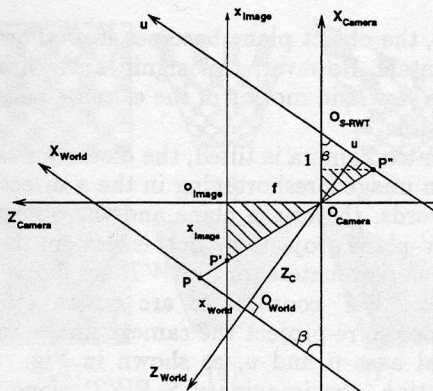


Figure 3: Projective model of perspective distortion compensation by the S-RWT (only the X - Z plane is shown)

3.2.1 Perspective distortion caused by camera tilt

A projective model for camera tilt is illustrated in Fig. 3. 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. 2. 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. Fig. 3 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 and the world X - Y planes. The object plane is set on the world X - Y plane. The translation Z_c between the camera origin and object plane is not affected by the 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 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 image of the world X - Z plane is projected on the image x - y plane. The projected image is obviously not the canonical frontal view or frontal view. It 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.

3.2.2 Perspective distortion compensation by S-RWT

The most straightforward way to remove perspective distortion is rotating the camera such that the image plane is parallel to the object plane. Conse-

quently, the object plane becomes frontal or canonical frontal. However, this significantly limits the possible view and motion of the camera, and is thus not feasible.

When the camera is tilted, the distortion is essentially an image foreshortening in the x direction. In other words, the image plane and the object plane form a V-plane projection. In this section, the transformation parameters for S -RWT are derived.

The S -RWT coordinates are on an imaginary plane used to re-project the camera image and with principal axes u and v , as shown in Fig. 3. 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 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. The re-projection of an image pixel onto the S -RWT plane through the camera optical center O_{camera} is a V-plane projection. The interception angle between them is β . By similar triangles, the re-projected image onto the S -RWT plane is a canonical frontal view of the object plane.

Suppose a point P on the object plane is projected onto the image plane at P' and re-projected onto the S -RWT plane at P'' . By carrying out the mapping between P' and P'' , the distorted image on x - y plane can be compensated. For the shaded area in Fig. 3, the triangles are similar. By equalizing the length proportions of the sides, we obtain:

$$\frac{-x}{f} = \frac{1 + u \cos \beta}{-u \sin \beta} \quad (7)$$

$$\frac{y}{f} = \frac{v}{u \sin \beta} \quad (8)$$

where x , y are the image coordinates and u , v are the S -RWT coordinates.

By rearranging Eqs (7) and (8), the transformation from image plane to S -RWT plane is,

$$u = \frac{\frac{f}{\sin \beta}}{x - \frac{f}{\tan \beta}} \quad (9)$$

$$v = \frac{y}{x - \frac{f}{\tan \beta}} \quad (10)$$

This latter transform is the *Shifted Reciprocal Wedge Transform* [2] with shift constant $a = -\frac{f}{\tan \beta}$. Since projection of the original camera image to the imaginary S -RWT plane forms a V-plane

²An additional constant scale factor $\frac{f}{\sin \beta}$ in Eq. (9) is needed in order to compensate the perspective distortion.

projection with the projection center at the camera optical center (see Fig. 3), this relation shows that the S -RWT can be implemented with the 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 .

4 GS-RWT for Compensating Distortion by Camera Pan-Tilt

For a point $[X, Y, Z_c]^T$ on the object plane, the coordinate after camera pan-tilt is

$$\begin{aligned} [X', Y', Z']^T &= T_{world}(\alpha, \beta) \cdot [X, Y, Z_c]^T \\ &= Rot_Y(-\beta) \cdot Rot_X(-\alpha) \cdot [X, Y, Z_c]^T \\ &= \begin{bmatrix} X \cos \beta + Y \sin \alpha - Z_c \cos \alpha \sin \beta \\ Y \cos \alpha + Z_c \sin \alpha \\ X \sin \beta - Y \sin \alpha \cos \beta + Z_c \cos \alpha \cos \beta \end{bmatrix} \quad (11) \end{aligned}$$

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

$$\begin{bmatrix} x_{image} \\ y_{image} \end{bmatrix} = \begin{bmatrix} f \frac{X \cos \beta + Y \sin \alpha - Z_c \cos \alpha \sin \beta}{X \sin \beta - Y \sin \alpha \cos \beta + Z_c \cos \alpha \cos \beta} \\ f \frac{Y \cos \alpha + Z_c \sin \alpha}{X \sin \beta - Y \sin \alpha \cos \beta + Z_c \cos \alpha \cos \beta} \end{bmatrix} \quad (12)$$

Eq. (12) represents the perspective distortion caused by the camera pan-tilt movements. We assume that the camera pose has five degrees of freedom: two for rotation and three for translation.

4.1 GS-RWT

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 frontal view of the object plane. Fig. 4 illustrates the object plane and camera image plane after the camera pan-tilt movements. Since the image plane is not parallel to the object plane, the projection of the object on the image plane is perspective distorted as represented in Eq. (12). Some variables introduced in Fig. 4 are 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 .

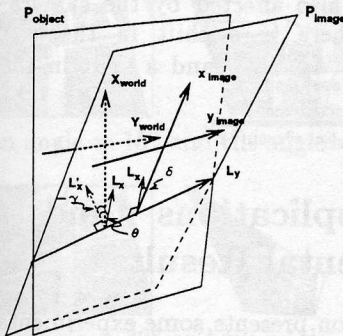


Figure 4: Configuration of image plane and object plane after camera pan-tilt

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.

In [4] the *GS-RWT* is developed to compensate the 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 *Generalized Shifted Reciprocal-Wedge Transform* is an extension of the *S-RWT*. The *GS-RWT* is a 2D transform performed on the image domain.

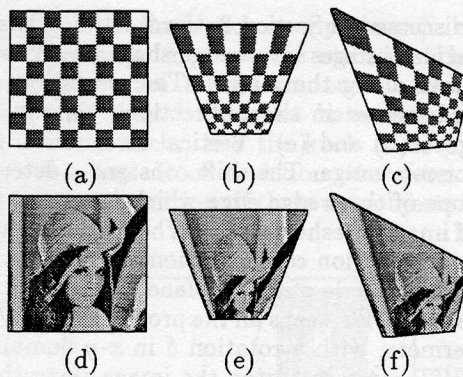


Figure 5: The Generalized Shifted Reciprocal Wedge Transform. (a) A rectangular grid. (b) The *S-RWT* image which foreshortens the image in vertical x -axis. (c) The *GS-RWT* image in which the direction of foreshortening deviates from principal axes. (d) (e) (f) The lady's image.

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. The first two steps restores a frontal view of the object, the third step obtains a canonical frontal view.

$$T_{GS-RWT}(\gamma, a, \delta) = Rot(-\gamma)T_{S-RWT}(a)Rot(\delta), \quad (13)$$

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

In [4] it is shown that $Rot(\delta)$, $Rot(-\gamma)$, and $T_{S-RWT}(a)$ are entirely determined by the pan and tilt angles α and β :

$$Rot(\delta) = \begin{bmatrix} \frac{\cos \alpha \sin \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & \frac{-\sin \alpha}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & 0 \\ \frac{\sin \alpha}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & \frac{\cos \alpha \sin \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (14)$$

$$Rot(-\gamma) = \begin{bmatrix} \frac{\sin \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & \frac{\sin \alpha \cos \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & 0 \\ \frac{-\sin \alpha \cos \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & \frac{\sin \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$\text{the constant } a = -\frac{f}{\tan \theta} = \frac{-f \cos \alpha \cos \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}},$$

$$T_{S-RWT}(a) = \begin{bmatrix} 0 & 0 & \frac{f}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} \\ 0 & 1 & 0 \\ 1 & 0 & \frac{-f \cos \alpha \cos \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} \end{bmatrix} \quad (16)$$

As discussed in Section 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. 5 (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 applying 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. 5 (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.

4.2 Perspective Distortion Compensation

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. (11). 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,

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = T_{GS-RWT}(\alpha, \beta) \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}$$

$$= \begin{bmatrix} \frac{X(\sin^2 \alpha \cos^2 \beta + \sin^2 \beta) + Z_c \cos \alpha \sin \beta \cos \beta}{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta} \\ \frac{Y(\sin^2 \alpha \cos^2 \beta + \sin^2 \beta) - Z_c \sin \alpha \cos \alpha \cos^2 \beta}{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta} \\ \frac{-Z_c}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} \end{bmatrix} \quad (17)$$

By Eq. (17), the 2D coordinate of the resulting image is calculated from $u_{image} = \frac{u}{w}$ and $v_{image} = \frac{v}{w}$,

$$\begin{bmatrix} u_{image} \\ v_{image} \end{bmatrix} = \begin{bmatrix} -X \cdot \frac{\sin^2 \alpha \cos^2 \beta + \sin^2 \beta}{Z_c \sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} - \frac{\cos \alpha \sin \beta \cos \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} \\ -Y \cdot \frac{\sin^2 \alpha \cos^2 \beta + \sin^2 \beta}{Z_c \sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} + \frac{\sin \alpha \cos \alpha \cos^2 \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}} \end{bmatrix} \quad (18)$$

From Eq. (18), 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 $\frac{\sin \alpha \cos \alpha \cos^2 \beta}{\sqrt{\sin^2 \alpha + \cos^2 \alpha \sin^2 \beta}}$.

5 Applications and Experimental Results

This section presents some experimental results. 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.

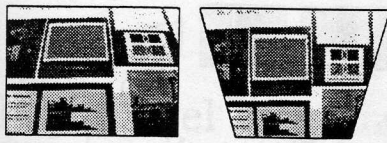
5.1 Perspective Distortion Compensation by GS-RWT

Corridor scenes taken by a CCD camera with different pan-tilt angles are shown in the left column of Fig. 6. 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. The *GS-RWT* parameters for the set of images in Fig. 6 are shown in Table 1. The constant reflecting the camera focal length was calibrated to be 280 in unit of image pixels.

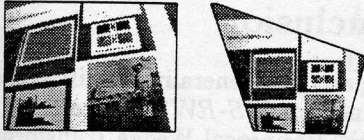
Table 1: Parameters of the *GS-RWT*

Items	Image a	Image b	Image c
Pan angle	0°	15°	30°
Tilt angle	30°	30°	30°
δ	0°	28°	49°
Main van. pt	(0, 485)	(-201, 376)	(-240, 207)
a	485	427	317
γ	0°	-24°	-41°

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 1 are



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



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



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

Figure 6: Perspective distortion compensation for different camera pan-tilt angles

expressed in terms of image pixels with the origin at the center of the camera image. Since the images in Fig. 6 were obtained with non-negative pan-tilt angles, the main vanishing points are located in the second quadrant. The magnitude of a indicates the rate of space-variant resolution.

5.2 GS-RWT for Visual Guidance on a Mobile Robot

5.2.1 System structure

The system is based on a wheeled mobile robot Nomad 200, a computer controlled pan-tilt unit with a CCD camera mounted on the top. The 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 direction of Sonar sensor 0 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. 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.

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.

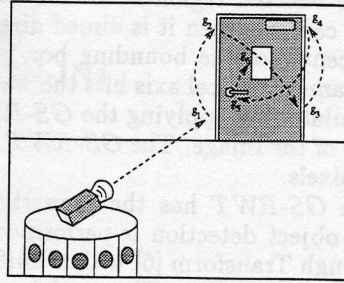


Figure 7: Gaze change for door detection

Gaze change is context-driven. It is based on the results of sub-object recognition. Fig. 7 illustrates a sequence of the gaze changes and Fig. 8 displays the camera image taken from six fixation points. In Fig. 8(a), a door-boundary-like vertical line is seen by the robot which is traveling down the corridor from the left to right. In Fig. 8(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. 8 (c) and (d). In the previous steps, the Hough Transform line detection algorithm [5] is used to find the door boundaries. Further checking is performed in Fig. 8(e) and (f) to detect the door knob and the peek window.

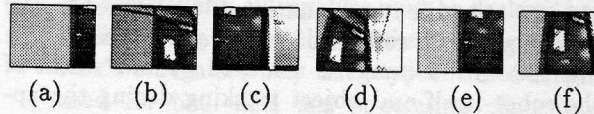


Figure 8: Images taken from the six fixation points for door detection

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.

The corresponding *GS-RWT* images for every probe point are shown in Fig. 9. Bounds are applied to retain only the central part of the image as

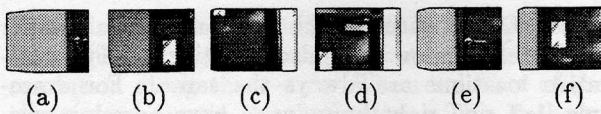


Figure 9: The *GS-RWT* images for the six fixation points

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.

Since the *GS-RWT* has the property of shape-preserving, object detection is performed using the General Hough Transform [6] with a uniform model on the *GS-RWT* image. The models of the door knob and peek window are generated off-line and used throughout the navigation task.

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. (18). 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. 9(b) are -15° and 30° , respectively; those for Fig. 9(d) are -35° and 26° , respectively.

5.2.3 Object tracking and approaching

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. The high resolution image is achieved by movement of the robot itself and object tracking during the approaching process.

Suppose the robot is directed to find a switch panel on the wall (See Fig. 10). 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 approximately 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 calculated 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.

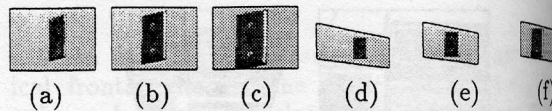


Figure 10: Object tracking and approaching

6 Conclusion

In this paper, the Generalized Shifted Reciprocal-Wedge Transform (*GS-RWT*) is presented as an extension to the Reciprocal-Wedge Transform (*RWT*). The *GS-RWT* can be represented in the matrix notation as a composite transform of one *S-RWT* and two rotations. The anisotropic variable resolution is derived from the *S-RWT* in which the shift constant a is used to compensate the perspective distortion. 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*. The space-variant resolution and shape-preserving properties of the *GS-RWT* are exploited in compensating perspective distortions caused by camera pan-tilt movements.

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. 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. The technique is applied to visual guidance of mobile robot navigation, it is shown to simplify the robot gaze control, 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 are facilitated since object recognition is performed on the recovered images in which the original shape of the object is preserved.

References

- [1] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford University Press, 1993.
- [2] 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.
- [3] F. Tong and Z.N. Li. Camera model for reciprocal-wedge transform. *Image and Vision Computing*, 14(5):339-351, 1996.
- [4] X.O. Ren. *Visual guidance for a mobile robot using the generalized shifted reciprocal-wedge transform*. M.Sc. thesis, Simon Fraser University, 1996.
- [5] 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.
- [6] D.H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111-122, 1981.