

# Reciprocal-Wedge Transform for Space-Variant Sensing

Frank Tong and Ze-Nian Li

**Abstract**—The Reciprocal-Wedge Transform (RWT) is presented as an alternative to the log-polar transform which has been a popular model for space-variant sensing in computer vision. The log-polar transform provides efficient data reduction. It simplifies the centric rotational and scaling image transformations. However, it adversely complicates the linear features and translational transformations. The RWT facilitates an anisotropic variable resolution. Unlike the log-polar, its variable resolution is predominantly in one dimension. Consequently, the RWT preserves linearity of lines and translations in the original image. In this paper, a concise matrix representation of the RWT is presented. Its properties in geometrical transformations and data reduction are described. A projective model for the transform and a potential hardware RWT camera design are also illustrated.

As examples of initial applications, the RWT is used for finding road directions in navigation, and for recovering depth in motion stereo. Two types of motion stereo are presented, namely the longitudinal and lateral motion stereo. In all cases, the RWT images offer much reduced and adequate data owing to the variable resolution. In road navigation, perspective distortion of the road image is readily corrected by the variable resolution of the RWT. In cases of the motion stereo, the correspondence problem in the RWT domain is reduced to a simpler problem of extracting collinear points in the epipolar plane. Preliminary experimental results from test images of road-vehicle navigation and moving objects on a miniature assembly line are demonstrated.

**Index Terms**—active vision, ego motion, motion stereo, navigation, reciprocal-wedge transform, space-variant sensing

## I. INTRODUCTION

ACTIVE vision has been advocated by many researchers [1], [2], [3], [4], [5]. It is argued that perception is not a passive, but an active process of exploratory, probing, and searching activities. An active visual system differs from a passive system in its purposive interaction with the world. Some interesting results in active vision include smart sensing using multiresolution images in the pyramid [6], fixation for 3D motion estimation [3], [7], active stereo using focus, vergence control [8], [9], and purposively adjusting multiple views for 3D object recognition [10], [11].

It has been argued that foveate sensors are central to the sensing mechanism of an active vision system because they are economical and effective when coupled with active control.

Research into anthropomorphic space-variant resolution sensors now receives much attention. The human visual system has a special saccadic behavior for quickly directing the focus of attention to different spatial targets [12], [13]. A foveate sensor coupled with fast and precise gaze control form the distinctive feature of the sensing mechanism of an active agent. In nature, human retina has a fovea which is a small region ( $1-2^\circ$ ) near the optical axis. The foveal resolution is compared to the peripheral resolution by orders of magnitude [13]. A design of this kind realizes an economical structure of sensor hardware supporting simultaneously a wide visual field and high acuity locally.

The study of Schwartz [14] shows that the cortical image of the retinal stimulus resembles a log-polar conformal mapping. Sandini and Tagliasco [15] argues that the retina sensor offers a good compromise among large visual field, acceptable resolution, and data reduction. The log-polar transformation is defined as  $w = (\log r, \theta)$  [16], where  $r$  and  $\theta$  are the polar coordinates of the original image. By exploiting the polar coordinates, it simplifies centric scaling and rotation [16], because these transformations now become shift operations in the  $\log r$  and  $\theta$  dimensions, respectively. As shown by Sandini and Dario [17], the scaling and centric rotational invariances of the log-polar transform make it a useful tool for 2D object recognition. The transform is also shown to be effective for estimation of time-to-impact from optical flow [18]. However, there is a major drawback with the log-polar transform. In general, image patterns of linear structures and translational movements are distorted into streamlines of log-sine curves [16] which adversely complicates the analysis of these common problems in computer vision.

At ICCV '93, we proposed a new transform, which is named *Reciprocal-Wedge Transform (RWT)* [19]. The RWT exhibits nice properties in computing geometric transformations owing to its concise matrix notation. As the log-polar, the RWT supports variable-resolution sensing and thus facilitates efficient data reduction. In particular, the variable resolution is anisotropic, predominantly in one dimension. Consequently, the RWT preserves linear features in the original image. That renders the transform especially suitable for vision algorithms that rely on linear structures, and vision problems that are translational in nature, e.g., line detection, stereo correspondence, linear motion, etc. For example, it is shown in [20] that vision systems for parts inspection in automated manufacturing and vehicle navigation in road driving will benefit from the anisotropic space-variant RWT representation.

A projective RWT model is presented in [19] which lends itself to a potential hardware implementation of the RWT

Manuscript received Nov. 3, 1993; revised Dec. 7, 1994.  
F. Tong is a member of the Computer Information Systems faculty at University College of the Fraser Valley, Burnaby, British Columbia, Canada.  
Z.N. Li is with the School of Computing Science at Simon Fraser University, Burnaby, British Columbia, Canada; e-mail: li@cs.sfu.ca.  
IEEECS Log Number P95051.

projection cameras. Unlike the log-polar sensor [21], [22], circular and variable sampling is not a requirement of the sensing unit. Hence, an ordinary sensor array of rectangular tessellation and uniform grid size can be used which is much cheaper to fabricate.

As one of the first applications of the RWT, a simple road navigation example is illustrated. It shows the perspective distortion of the road image is readily corrected by the variable resolution of the RWT, which enables a more efficient search of the road direction on the reduced data.

The RWT is also shown to be applicable to *stereo vision* for depth recovery. One of the difficult problems in stereo vision is *correspondence* [23]. Once corresponding points in the pair of images are identified, their *disparity* values can be calculated and used to recover the depth. The last part of this paper shows the application of the RWT to the correspondence process in *motion stereo* [24]. Two types of motion stereo are discussed, namely the longitudinal and lateral motion stereo. In both cases, the properties of the anisotropic variable resolution and linear features in the RWT domain are exploited to yield efficient variable-resolution algorithms which work on the much reduced image data. The difficult and computationally expensive correspondence problem in both motion stereo cases is effectively reduced to an easier problem of finding collinear points in the epipolar planes, which is later solved by a voting algorithm for accumulating multiple evidence.

Section II introduces the transform and its properties. Section III delineates the projective model and its potential hardware implementation. Section IV describes application of the RWT in road navigation and some initial result using a synthetic image. Applications of the RWT in two motion stereo cases and preliminary test results using real-world images are discussed in Section V. Section VI concludes the paper.

## I. THE RECIPROCAL-WEDGE TRANSFORM

The Reciprocal-Wedge Transform (RWT) is defined as a mapping from  $x$ - $y$  to  $u$ - $v$  such that

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

In practice, the singularity at  $x = 0$  (the center strip) can be excluded from the range of the variable-resolution processing. In case that the center strip is included, a shift parameter  $a$  can be added to  $x$  to shift it away from zero. Let us call this variant formulation the *Shifted Reciprocal-Wedge Transform (S-RWT)*.

$$u = 1/(x+a), v = y/(x+a) \quad (2)$$

The shift parameter is also included in the log-polar transform in the similar manner to take care of the singularity of the logarithmic function [14], [25]. The impact of the parameter  $a$  is limited to a horizontal shift of the  $x$  axis in the Cartesian image. In this paper, the formulas in (1) are used for simplicity in the analyses. Discussion will be extended to the S-RWT, when necessary.

### A. Matrix Representation

A concise representation for the transformation is derivable using the matrix notation. Adopting the homogeneous coordi-

nates, the RWT defined in (1) can be formulated as a cross-diagonal matrix of 1s, and the transformation can be computed as matrix operations.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{T}^{-1}, \quad (3)$$

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

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

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

The sign “ $\approx$ ” means equality within the homogeneous coordinate representation.

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

The lady image in Fig. 1(a) is used to illustrate how the Cartesian coordinates are mapped to the RWT domain back and forth. The transformed image in Fig. 1(b) shows a wedge-shape in an inside-out fashion because of the scaling effect of the  $x$  reciprocal. Note the blurring at the periphery of Fig. 1(c). In Fig. 1(d-f), the grid image is used to demonstrate the variable resolution of the transform. It is the differential magnification ratio across the width of the image that facilitates the continuously changing scale of image resolution from the center to the periphery.

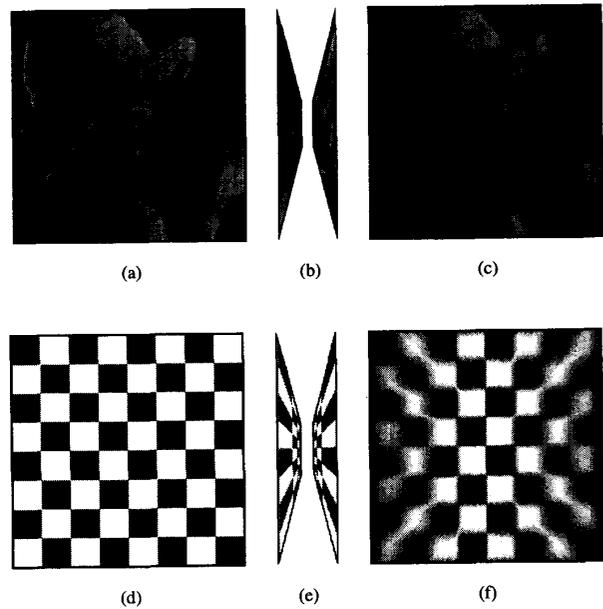


Fig. 1. The Reciprocal-Wedge Transform. (a) The lady image. (b) The RWT image shows two inside-out wedges. (c) The image when transformed back to the Cartesian domain. (d) A rectangular grid. (e) The RWT image. (f) The grid transformed back to illustrate the resolution varying from the center to the periphery.

The S-RWT from (2) was used in the transformation of Fig. 1 in order to cover the whole image including the singularity of the  $x$  reciprocal in the center region as well. Both the forward and backward operators for the S-RWT remain the same cross-diagonal matrix (3) except the additional parameter  $a$ .

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

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

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

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

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

Since both  $\mathbf{T}$  and  $\mathbf{T}^{-1}$  are cross-diagonal matrices of 1s (3), their effect on  $\mathbf{M}$  involves only row and column interchange. Thus, the respective matrices for the RWT domain can easily be derived as:

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

## B. Transformation on Linear Structures

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

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

### B.1 Preservation of Linear Features

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

$$L_{xy}: y = m \cdot x + c, \quad L_{uv}: v = c \cdot u + m. \quad (4)$$

Given  $L_{xy}$ , the equation for  $L_{uv}$  is readily obtained by substi-

tuting  $x$  and  $y$  in  $L_{xy}$  with  $1/u$  and  $v/u$ , respectively. It is obvious that the transformed structure  $L_{uv}$  is also a line, which implies that the linearity of the line is preserved.<sup>1</sup> It is interesting to note that the values for the slope and intercept are interchanged between the transformation dual. Inferring from that, parallel lines with the same slope in  $x$ - $y$  will be mapped to  $u$ - $v$  lines concurrent at the same  $v$ -intercept. Inversely, lines concurrent at the same  $y$ -intercept will form parallel lines in the  $u$ - $v$  domain.

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

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

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

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

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

### B.2. Line Detection Using the Hough Transform

The Hough transform [26] provides a powerful tool for feature detection. The technique is most effective for line detection. The preservation of the linearity of lines over the RWT implies that line detection using the Hough transform would be as simple in the RWT as in the Cartesian domain. With the switching between the slope and intercept parameters (4), the vote patterns in the Hough space for the Cartesian and the RWT images form an interesting dual of reflection about the main diagonal of the Hough space. (See Fig. 2(c).)

## C. Anisotropic Space-Variant Resolution

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

The essence of the RWT is the introduction of the reciprocity. The variable resolution is primarily embedded in the  $x$  dimension. It yields a grid whose resolution is variable for different  $x$ s, but uniform along the  $y$  dimension for any fixed  $x$ . The result is an anisotropic space-variant resolution, which is evident from the wedge-shaped grid in Fig. 1(e).

The anisotropy can also be inferred from the partial derivatives of the RWT expressions from (1):

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

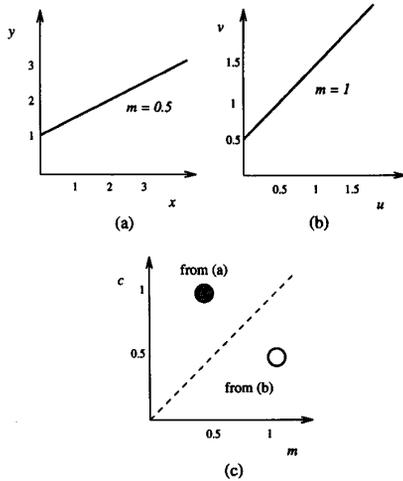


Fig. 2. The duality relationship of linear structures in the RWT. (a) A line in the  $x$ - $y$  domain with a slope 0.5 and the intercept 1. (b) The dual in the  $u$ - $v$  domain. The slope is 1 and the intercept is 0.5, inversely. (c) The Hough space showing the peaks from (a) and (b), respectively. They form a reflection about the main diagonal.

$$\|\partial(u, v)/\partial x\| = \|\partial u/\partial x, \partial v/\partial x\| = \sqrt{(1+y^2)}/x^2, \quad (5)$$

$$\|\partial(u, v)/\partial y\| = \|\partial u/\partial y, \partial v/\partial y\| = 1/x, \quad (6)$$

where  $\|\cdot\|$  denotes the vector norm. Equations (5) and (6) show that the pixel resolution does not vary in the same scale for different directions. The grid width in the  $x$  direction (for a fixed  $y$ ) is mapped to a size diminishing in reciprocal of  $x^2$ . In the  $y$  direction, the grid height is mapped by a function of  $1/x$  to a uniform size independent of the  $y$  value. Furthermore,

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

Hence, the absolute value of the above Jacobian determinant is

$$|\partial(u, v)/\partial(x, y)| = 1/x^3 \quad (7)$$

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

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

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

The area of a pixel thus diminishes isotropically in the rate of  $1/r^2$ . The isotropic variable resolution of the log-polar transform is related to the conformality of the log-polar mapping [16].

### III. A PROJECTIVE MODEL AND ITS HARDWARE IMPLEMENTATION

#### A. The RWT Projective Model

Fig. 1(d) appears like the view of a picture from a grazing angle. In fact, one could regard the RWT as an oblique perspective projection of an image on a plane perpendicular to it. Examine the perspective projection in which the three-dimensional  $XYZ$  space is projected onto the two-dimensional  $Z$ - $Y$  plane at  $X = 1$  (Fig. 3). Let the three-dimensional point be  $(X, Y, Z)$  and the projection be  $(Z', Y')$ .

$$Z' = Z/X, \quad Y' = Y/X. \quad (8)$$

Now, the equations in (1) can be made equivalent to those in (8) if the terms  $x, y, 1, u, v$  in (1) are unified with the  $X, Y, Z, Z', Y'$  in (8), respectively. In that sense, the RWT described by (1) can also be viewed as a perspective reprojection in which the original image is now placed on the  $X$ - $Y$  plane at  $Z = 1$ , and it is projected onto the  $Z$ - $Y$  plane at  $X = 1$ .<sup>2</sup>

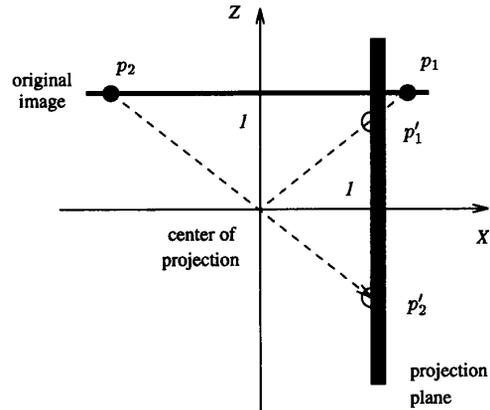


Fig. 3. A perspective projection model. The original image is placed on the  $X$ - $Y$  plane at  $Z = 1$ . It is projected onto the  $Z$ - $Y$  projection plane at  $X = 1$ . The pixels  $p_1$  and  $p_2$  are projected to  $p_1'$  and  $p_2'$  respectively.

#### B. A Potential Hardware Implementation

It is interesting to notice that the RWT can potentially be implemented in hardware. Since the RWT image can be considered as another perspective projection of the Cartesian image onto an orthogonal projection plane, we can cascade the two processes into one. Fig. 4 illustrates the idea. The sensor is fitted directly on the RWT projection plane mounted sideways. Thereby, the rays from the imaged objects strike directly onto the RWT sensor plane. The sensor plane is installed in two half-planes, the left and right ones, respectively, for the convenience of taking care of objects on each side of the optical axis.

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

The RWT camera can use a uniform sensor, which is cheap to fabricate. Space-variant sensing is realized by the oblique projection on the sensor plane. The same sensor area on the plane yields variable area coverage of the visual field depending on the angle of projection. In fact, one can also alter the position, orientation or even the shape of the sensor plane to produce different space-variant sensors.

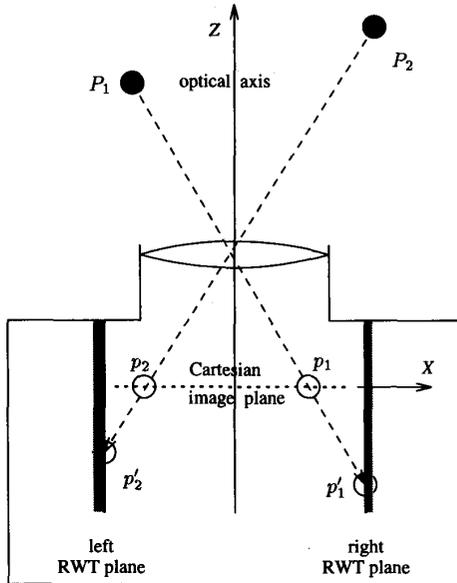


Fig. 4. An RWT projection camera. The camera has its sensor placed on the left and right horizontal planes. Light ( $P_1, P_2$ ) passing through the lens forms a virtual image ( $p_1, p_2$ ) on the frontal focal plane which is then further projected onto the sideways-positioned RWT sensor planes ( $p'_1, p'_2$ ).

The above discussion presents the rudimentary idea of the RWT camera design. A prominent problem is the requirement of focusing on a deep image plane in parallel to the optical axis. In our recent paper [27], a new hardware camera model is proposed which overcomes the focus problem by using a lens focusing the paraxial image onto an orthogonally placed RWT plane.

#### IV. APPLICATION OF RWT IN ROAD NAVIGATION

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

Different approaches have been devised for road detection in various experiments of autonomous land vehicles. An area-based voting scheme based on the Hough transform is applied to compute the direction of the road in the CMU Navlab [28]. In the VaMoRs project [29], visual features of the road edges are detected based on the "Gestalt" hypothesis under adverse situations of shadows and absent lane-markings.

Both methods search over the perspective images for road features. The drawback is that the nearby section of the road gets overly attended whereas the far side toward the horizon is

disproportionately under-sampled. Arguably, this differential scale of details is not suitable for driving on the road. Normally, one would pay more attention to a reasonable distance to see the general direction of the road, while at the same time be aware of the road segment right in front.

Lotufo et al. [30] present the *plan-view transformation* method for road navigation. The original perspective road

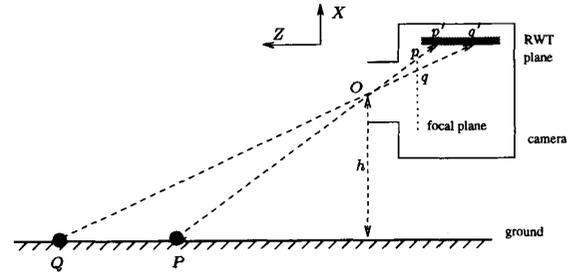


Fig. 5. Perspective inversion effected by the RWT projection. The image size ( $p-q$ ) varies with the position of the segment  $PQ$ , whereas the size of the RWT image ( $p'-q'$ ) does not.

image is projected to a grid inclined by a pan angle  $\theta$ , which is chosen so that the road edges are nearly parallel to the boundaries of the grid. It is also reported that the new images are of a reduced size (typically by a factor of 32).

Effectively, the RWT re-samples the image to a variable resolution which counterbalances the differential scale of details in the perspective projection. Fig. 5 illustrates the perspective inversion. As the space-variant resolution is defined along the  $x$  dimension in the RWT model, the vertical axis is denoted as the  $X$ -axis of the world coordinate system shown in Fig. 5.

With the road segment  $PQ$  projected onto a horizontal plane in the RWT camera, the projection ( $p', q'$ ) does not change in magnification, and the perspective projection is practically inverted.

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

The road in the RWT image appears as though it were from an aerial view (Fig. 7(b)). However, an important difference is that the RWT camera is pointing toward the front which is vital to driving. Moreover, the nearby section is sampled at a much reduced resolution. The overall data volume can be greatly reduced to achieve a comparable performance.

In the CMU Navlab project [28], a road is perceived as converging at one point on the vanishing line, and is parameterized by  $P$  and  $\theta$  (Fig. 6). As discussed in Section II-B.2, the detection of converging lines at the vanishing point for finding the road direction can be carried out by detecting parallel road boundaries in the RWT image. The technique of the Hough transform is equally applicable.

When edge tracking is employed to calculate the geometric model of the road in the VaMoRs project, the RWT image benefits by eliminating the variable search ranges for road features in the near and distant sections of the road. In all cases, the RWT sup-

ports an efficient representation of the road image as the data volume is greatly reduced by its spatially variable sampling.

### A. Experimental Result for Finding Road Direction

Fig. 7(a) is a synthetic image of a road scene. The image has a resolution of  $128 \times 256$ . Fig. 7(b) is its RWT image of the size  $32 \times 128$ . The remote section of the road has retained its resolution whereas the excessive information at the near side is suppressed. The total area of the search region is significantly reduced. The direction of the road is detected using the Hough method described in [28]. In the RWT image, the road direction is detected as  $\tan \phi = -22$ , yielding the position of the vanishing point  $P$  in the original Cartesian image.

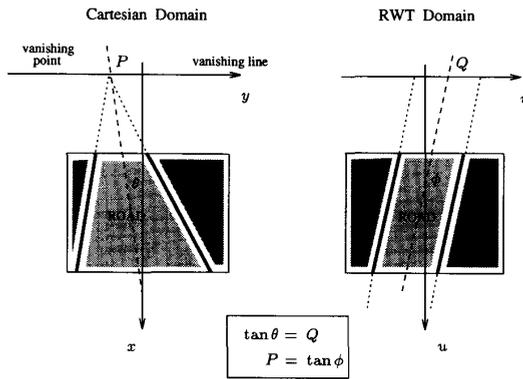


Fig. 6. The RWT dual of the road image. The vanishing point in the Cartesian domain just becomes the direction in the RWT domain and vice versa.

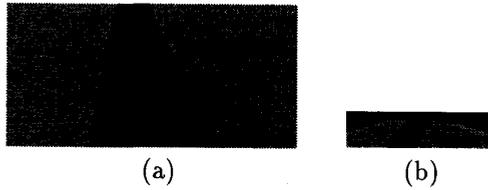


Fig. 7. The synthetic image of a road scene. (a) The road image of resolution  $128 \times 256$ . (b) The RWT image of size  $32 \times 128$ .

## V. APPLICATION OF RWT IN MOTION STEREO

Okutomi and Kanade pointed out in [31], the distance between the pair of cameras in stereo vision greatly affects the precision and error rate of the correspondence process. A short baseline will provide less precision whereas a longer baseline will result in higher error rate due to false match. To alleviate the dilemma, they proposed the *multiple-baseline stereo* method [31] where different baselines are generated by lateral displacements of a camera.

Consider a manufacturing environment with intelligent robots working on assembly lines, where the belts are moving at a relatively constant speed. Multiple snapshots of the moving

objects on the belt can be taken in a rapid succession by a single camera. The controlled belt movement provides the necessary stereo disparity. Moreover, it can guarantee that the disparity occurs only along the epipolar lines. This method is called *Motion Stereo* [24]. Its greatest advantage is the simplicity in camera control and calibration. Suppose the camera is looking down the  $Z$  direction, i.e., its optical axis is the  $Z$ -axis. We call the above moving belt situation *lateral motion stereo* where objects move on a  $Z = Z_0$  plane, perpendicular to the  $Z$ -axis. Another type is *longitudinal motion stereo* in which objects move along the  $Z$  direction, an autonomous vehicle traveling on the highway provides such an example.

Bolles, Baker, and Marimont [32] proposed a technique of epipolar-plane image analysis for determining structure from motion. It was pointed out that for straight-line and constant-speed camera motions, simple linear structures will be formed on the epipolar-planes (Fig. 8), where the slope of these lines indicates the depth of the feature points.

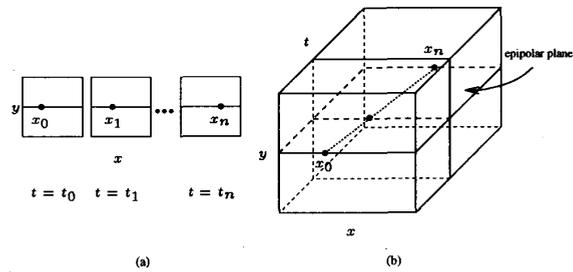


Fig. 8. Epipolar-plane image analysis. (a) A feature point moves along the epipolar line in the  $x$ - $y$  plane with a constant speed. (b) A linear locus is formed on the epipolar ( $x$ - $t$ ) plane in the  $xyt$  space.

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

### A. Longitudinal Motion Stereo

Depth recovery in autonomous vehicle navigation provides an example for the longitudinal motion stereo, in which the relative object movement is along the  $Z$  direction at a constant speed. Fig. 9(a) illustrates a point moving from position  $P_0(X_0, Y_0, Z_0)$  at  $t_0$  to position  $P_1(X_0, Y_0, Z_1)$  at  $t_1$ . The  $x$ -coordinates of its projections on the ordinary  $x$ - $y$  image plane are  $x_0$  and  $x_1$ . The corresponding images on the RWT  $u$ - $v$  plane are  $u_0$  and  $u_1$ . As shown, the focal lengths are  $f$  and  $f'$ , respectively. For simplicity (and with the deviation of a constant factor), it is assumed that  $f = f' = 1$ .

From similar triangles,

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

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

$$\frac{dx}{dt} = -\frac{X_0}{Z^2} \frac{dZ}{dt} = -\frac{CX_0}{Z^2} \propto \frac{1}{Z^2}, \quad (9)$$

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

If multiple images of the longitudinal motion stereo are used, then  $x_0^k, x_1^k, \dots, x_n^k$  are a sequence of corresponding points for the point  $P^k$  at  $t = t_0, t = t_1, \dots, t = t_n$  in the  $x-t$  epipolar plane. As shown in Fig. 9(b), their locus is nonlinear (a curve), which is implied by (9).

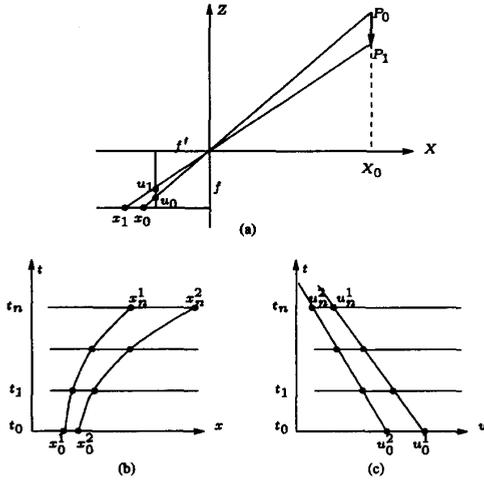


Fig. 9. (a) Longitudinal Motion Stereo. (b) The  $x-t$  plane from ordinary longitudinal motion stereo images. (c) The  $u-t$  plane after the RWT.

Now, it can be shown that the reciprocity of the RWT happens to counterbalance the above nonlinearity. From Fig. 9(a)

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

It follows,

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

or

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

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

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

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

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

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

#### A.1 Extension to Ego Motion

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

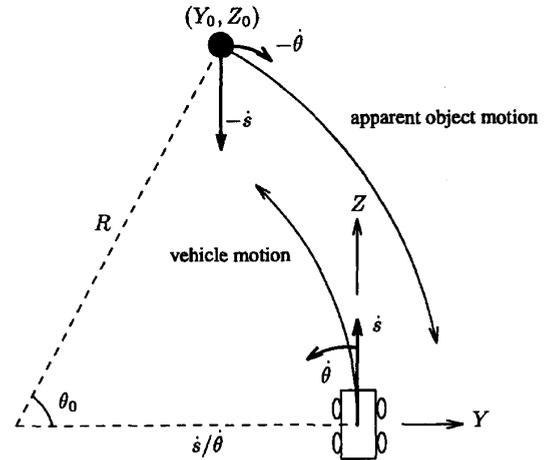


Fig. 10. Motion of an object in relation to the vehicle. In the world coordinates, the vehicle is traveling at an axial velocity of  $\dot{s}$  and rotational speed of  $\dot{\theta}$ , describing a circular path of radius  $\dot{s}/\dot{\theta}$ . In the viewer-centered coordinates of the vehicle driver, the object is moving in the opposite direction with the same speed. It also appears to move on a circular trajectory at the same center.

Assuming the vehicle is moving in an otherwise static world, the apparent motion of the world in the view of the vehicle driver is a composite motion of axial translation  $-\dot{s}$  and centric rotation  $-\dot{\theta}$ . Take the vehicle driver as the center of reference and align the  $Z$  axis with the direction of travel as depicted in Fig. 10. At time  $t$ , the position of the object is at  $(X, Y, Z)$ .

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

Solving the differential equations,

$$Y = R \cos(\theta_0 - \dot{\theta}t) - \dot{s}/\dot{\theta}, \quad (10)$$

$$Z = R \sin(\theta_0 - \dot{\theta}t). \quad (11)$$

The form of the equations indicates the circular path for the object's apparent motion.  $R$  is the radius of the circular path and the center is at  $Y = -\dot{s}/\dot{\theta}, Z = 0$ . At  $t = t_0 = 0$ , the object is at

the position  $(Y_0, Z_0)$ , which is  $Y_0 = R \cos \theta_0 - \dot{s}/\dot{\theta}$ ,  $Z_0 = R \sin \theta_0$ . Hence,  $\theta_0$  is the arc angle on the circular path at which the object is initially located.

From Fig. 9(a), the mapping from  $(Y, Z)$  to  $(u, v)$  is

$$u/f' = Z/X, \quad v/f' = Y/X. \quad (12)$$

Apply the mapping on (10) - (11). The image motion on the  $u$ - $v$  plane now is

$$u = \frac{f'}{X} R \sin(\theta_0 - \dot{\theta}t),$$

$$v = \frac{f'}{X} R \sin(\theta_0 - \dot{\theta}t) - \frac{f'}{X} \frac{\dot{s}}{\dot{\theta}}.$$

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

$$u = r \sin(\theta_0 - \dot{\theta}t), \quad (13)$$

$$v = r \cos(\theta_0 - \dot{\theta}t) - a. \quad (14)$$

Apparently, the  $u$ - $v$  motion is along a circular trajectory with the radius  $r$ , and the center of curvature is at  $u = 0$  and  $v = -a$  (see Fig. 11(a)).

Use  $\omega$  for the arc distance measured from the  $v$  axis along the circular trajectory as shown in Fig. 11(a). The advantage of using  $\omega$  is that it shows a linear relationship with  $t$  (see Fig. 11(b)).

$$\omega = r(\theta_0 - \dot{\theta}t). \quad (15)$$

One useful property of using the  $\omega$ - $t$  line is the readily computable extrapolated  $t$ -intercept. Putting  $\omega = 0$  in (15), the  $t$  indicates the time in which a point comes to the  $v$  axis. This time measure yields the *time-to-contact*.

$$t = \theta_0 / \dot{\theta}. \quad (16)$$

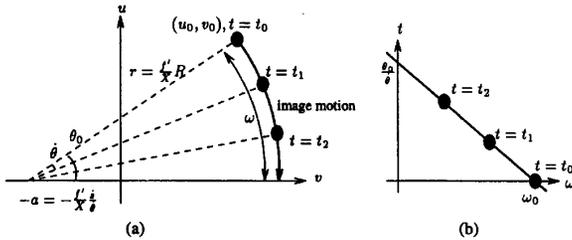


Fig. 11. Image motion in  $u$ - $v$ . (a) The image motion is a circular arc centered at  $-a = -f'/X \cdot \dot{s}/\dot{\theta}$  and has a radius of  $r = f'/X \cdot R$ . The initial arc angle is  $\theta_0$  for the point position  $(u_0, v_0)$  at  $t = t_0 = 0$ . The point is approaching at an angular speed  $\dot{\theta}$ . (b) When the arc length  $\omega$  is measured against  $t$ , it shows a linear relationship. The  $t$ -intercept is  $\theta_0/\dot{\theta}$ .

The  $x$ - $y$  uniform resolution image represents the perspective projection of the driving scene. The reciprocity of the RWT counterbalances the perspective nonlinearity and yields a linear mapping of the road surface (12). The linear mapping enables the preservation of the circular image motion as corresponding to the original vehicle motion. Such is not the case in

the  $x$ - $y$  image because of the perspective distortion

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

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

A search algorithm can be devised to find the circular trajectories on the  $u$ - $v$  plane as described in (13) - (14). When visualized in the 3D  $uv$  space, the circular trajectory becomes a helical curve. The search is essentially a problem of fitting the helical model to the  $uv$  data. Nevertheless, the search space is much restricted by exploiting the constraints due to the simple vehicle motion. The helical trajectory in  $uv$  has no more than two degrees of freedom even if none of the constants  $f'$ ,  $\dot{s}$ ,  $\dot{\theta}$  are known a priori. From (13) - (14), the center of the helix is on the  $v$  axis. Choose  $a$  for the position of the center in (14), the radius  $r$  and the arc length  $\omega$  for each feature point  $(u, v, t)$  from the RWT image sequence can be determined from (13) - (15). The helical trajectory of the point  $(u_0, v_0)$  corresponds to a straight line passing through  $\omega_0$  in the  $\omega$ - $t$  projection (Fig. 11). Now, choose a value for the line slope such that the line passing through  $\omega_0$  would fit to the  $\omega$ - $t$  projection of the feature points. The best fitted line over different values of  $a$  yields the best solution to the helical trajectory of  $(u_0, v_0)$ . By (16), the  $t$ -intercept of the  $\omega$ - $t$  indicates the time-to-contact.

The model of the longitudinal motion stereo for linear vehicle motion is a special case of this general model for ego motion. When  $\dot{\theta} \rightarrow 0$ ,  $\dot{s}/\dot{\theta}$  approaches  $\infty$  and so do  $r$  and  $a$  in (13) - (14). The circular trajectory therefore approaches a line along the  $u$  direction, and the arc length  $\omega$  now directly corresponds to the  $u$  coordinate. In the general model, the  $t$ -intercept in  $\omega$ - $t$  (16) indicates the time-to-contact for an object if the vehicle motion were to prevail. In the special case of linear vehicle motion, conveniently, the time-to-contact gives a direct measure of the depth of an object as well.

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

## B. Lateral Motion Stereo

This section uses the example of the moving assembly line mentioned earlier. For simplicity, let's first assume that the belt moves in the  $X$  dimension in the 3D space. Its projected movement on the  $x$ - $y$  plane is therefore along the  $x$  direction only. For a point  $x^k$  at  $y = y'$ ,  $x_0^k, x_1^k, \dots$ , and  $x_n^k$  are a sequence of corresponding points at  $t = t_0, t = t_1, \dots$ , and  $t = t_n$  in the  $x$ - $t$  epipolar plane from the original (ordinary) lateral motion stereo images, where the epipolar lines are horizontal (Fig. 12(a)). When the speed of the belt is constant and images are taken at equal intervals,  $x_0^k, x_1^k, \dots$ , and  $x_n^k$  fall on a single line in the  $x$ - $t$  plane, and  $\frac{dx}{dt} \propto \text{disparity } d$ . Hence, the correspondence

problem in the lateral motion stereo is equivalent to a problem of finding collinear points in the  $x$ - $t$  epipolar plane. Since the disparity is inversely proportional to the actual depth in the 3D

scene, it follows that  $\frac{dt}{dz} \propto \text{depth}$  of the point  $x^k$ .

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

$$\omega = \sqrt{u^2 + v^2} = \sqrt{1 + (y^l)^2} / x. \quad (17)$$

The epipolar plane for the lateral motion stereo becomes the  $\omega$ - $t$  plane as shown in Fig. 12(b). Note the new sequence of the corresponding points  $\omega_0^k, \omega_1^k, \dots$ , and  $\omega_n^k$  do not generally fall on a single line in the  $\omega$ - $t$  plane.

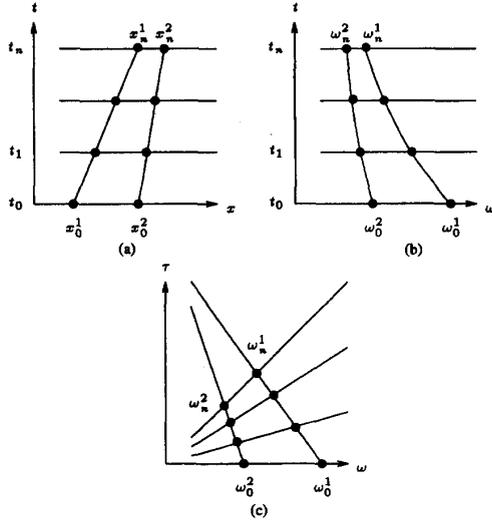


Fig. 12. Two sequences of edge points on epipolar planes from motion stereo images. (a) The  $x$ - $t$  plane from ordinary lateral motion stereo images. (b) The  $\omega$ - $t$  plane where  $\omega = \sqrt{1 + (y^l)^2} / x$ . (c) The  $\omega$ - $\tau$  plane where  $\omega = \sqrt{1 + (y^l)^2} / x$ , and  $\tau = t/x$ .

### B.1 Creation of a New $\omega$ - $\tau$ Epipolar Plane

To recover the linearity, an additional transformation

$$\tau = t / x \quad (18)$$

can be applied to the variable  $t$  which is similar to what is applied to  $y$  in the RWT. The  $x$ - $t$  epipolar plane from lateral motion stereo images is now converted into a new  $\omega$ - $\tau$  epipolar plane of the RWT images (Fig. 12(c)). The horizontal epipolar lines in the  $x$ - $t$  plane become concurrent epipolar lines merging toward the origin in the  $\omega$ - $\tau$  plane. The lines that connect the corresponding points now remain to be lines as well.

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

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

its transformation dual (derivable from (17), (18)) in the  $\omega$ - $\tau$  plane is  $L_{\omega\tau}$ :

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

The slope  $m'$  of line  $L_{xt}$  becomes the  $\tau$ -intercept of  $L_{\omega\tau}$  in the RWT motion stereo.<sup>3</sup> Instead of  $\frac{dt}{dx} \propto \text{depth}$ , it is now the case that the  $\tau$ -intercept of the line that connects the corresponding points in the  $\omega$ - $\tau$  plane is  $\propto \text{depth}$  of the point  $\omega^k$ .

### B.2 Extension to Any Linear Motion on $Z = Z_0$ Plane

Although it was assumed above that the belt moves along the  $X$  dimension only, this can be relaxed to any linear movement on a  $Z = Z_0$  plane in the 3D scene. The projected locus on the  $x$ - $y$  plane is the epipolar line  $L_{xy}$  of which the slope  $m$  and  $y$ -intercept  $c$  are known parameters.

As shown by (4), after the RWT the line  $L_{xy}$  is transformed to the new line  $L_{\omega\zeta}$ . Let  $\zeta$  be the length of the line segment  $L_{xy}$  from the  $y$ -axis to  $(x, y)$ ,

$$\zeta = \sqrt{x^2 + (y - c)^2} = \sqrt{1 + m^2} x. \quad (19)$$

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

$$L_{\zeta t} : t = m' \cdot \zeta + c'. \quad (20)$$

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

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

Because  $u = 1/x$ , use (19),

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

If we introduce a new parameter  $\tau = t / \zeta$ , then the line in the  $\zeta$ - $t$  plane (20) will be converted into a line in the  $\omega$ - $\tau$  plane,

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

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

### C. Search in the Epipolar Plane

As described above, the correspondence problem in both the longitudinal and lateral motion stereo can be reduced to a problem of searching for collinear points in the epipolar planes ( $u$ - $t$  plane for the longitudinal,  $\omega$ - $\tau$  for the lateral). Similar to the procedures for the Hough transform [26], a voting algorithm for accumulating multiple evidence can be developed. Without loss of generality, the search on the  $u$ - $t$  plane in the longitudinal stereo for linear motion will be used here to illustrate the method. (The extension to circular ego motion re-

3. This is similar to the transformation dual in (4), except the slope of  $L_{\omega\tau}$  is not  $c'$  because of the additional transformation on  $t$  (18).

quires a somewhat different search, i.e., search for helical curves in the  $uv$  space. By introducing  $\omega$  as the arc length, the problem was shown in Section V-A.1 to be equivalent to finding collinear points on the  $\omega$ - $t$  plane. For efficiency, a slightly different search algorithm was suggested earlier in Section V-A.1.)

In general, any point at  $t = t_i$  can be paired with any point at  $t = t_j$  ( $j > i$ ) to form a hypothetical line segment. Its intercept on the  $t$ -axis suggests a possible depth value which is inversely proportional to the disparity  $d$ . A 3D  $uvd$  voting space is created<sup>4</sup> and each hypothetical line will cast a vote at the position  $(u, v, d)$  in the  $uvd$  space. Since  $n + 1$  collinear points can form  $O(n^2)$  hypothetical lines and they will vote to the same  $(u, v, d)$ , a peak will be formed in the  $uvd$  space which indicates the consensus on the correct disparity value for the point  $(u, v)$ . The line detection problem can thus be solved by this voting procedure followed by a peak-detection procedure.

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

- **Use relatively long hypothetical voting lines.** Due to limitation of the image resolution there is always certain error in the  $u$ - $v$  coordinates, especially at the periphery of the RWT images. If the short hypothetical lines were to be used for voting, a small amount of error in the  $u$ - $v$  coordinates would result in relatively large errors in the calculation of the slope and intercept, and consequently the disparity values. A minimum length is therefore chosen to exclude the short voting lines.
- **Specify a reasonable range for depth.** A range of concerned depth can be represented as  $[T_{min}, T_{max}]$  to reduce the number of candidate pairs. The vertices  $T_{min}$  and  $T_{max}$  on the  $t$ -axis and the lower end point  $u_i^k$  form a triangle which defines the search region for the possible pairing end point  $u_j$ .

## D. Experimental Results for Motion Stereo

### D.1 Longitudinal Motion Stereo

A vehicle navigation example is used to illustrate the longitudinal motion stereo. Fig. 13(a) shows a CMU image sequence of a road scene obtained from a driving expedition. Four frames of an 8-snapshot sequence (each has a size of  $512 \times 512$  pixels) are shown to visualize the forward motion from driving. The RWT images of the motion sequence have been generated in software. The data reduction factor is over 90%. Fig. 13(b) shows the RWT edge images of the  $x$ - $y$  sequence.

Some implementation details should be dealt with when generating the RWT images. First, the  $X$ -axis in the world coordinate system is the vertical axis as indicated in Fig. 5. Accordingly, the  $x$ -axis in the  $x$ - $y$  images and the  $u$ -axis in the RWT ( $u$ - $v$ ) images are the vertical axes in these images. Sec-

ond, the model of our longitudinal stereo requires both the camera movement and its optical axis be along the  $Z$ -axis. According to this simple model, the FOE (Focus of Expansion) is always at the center of the  $x$ - $y$  road images. When dealing with FOE which is significantly off center because of intentional pan/tilt on the camera orientation, the FOE must be determined and used as the origin of the  $x$ - $y$  space for the RWT. This is the

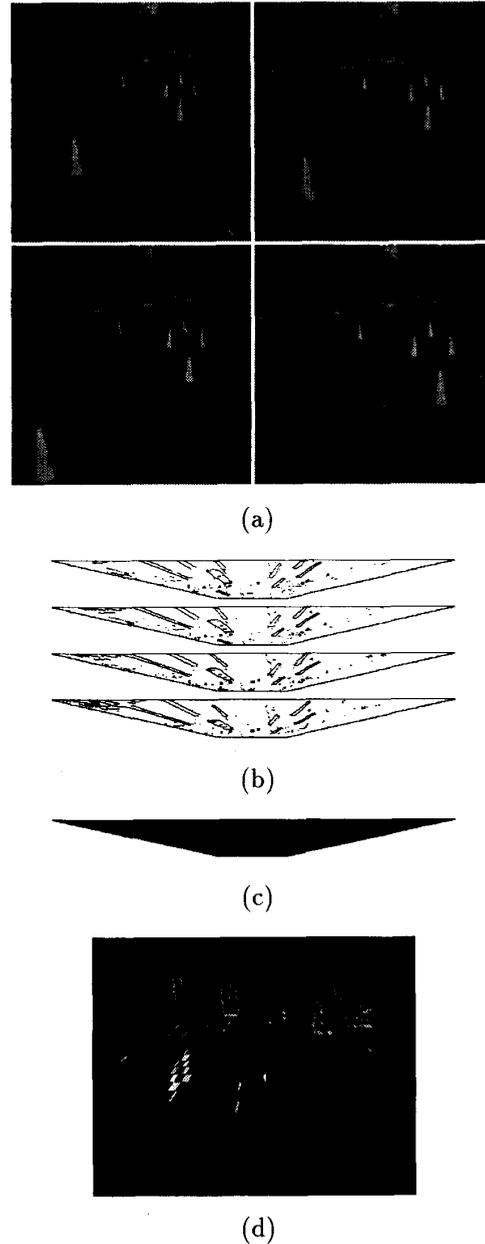


Fig. 13. Depth computation using the RWT in linear motion. (a) A sequence of a driving scene, only Image 1, 3, 5, and 8 are shown. (b) Edge images from the above RWT images. (c) Gray-level coded depth map computed from all eight images. (d) The depth map transformed back to the  $x$ - $y$  space (uniform-resolution) for visual apprehension.

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

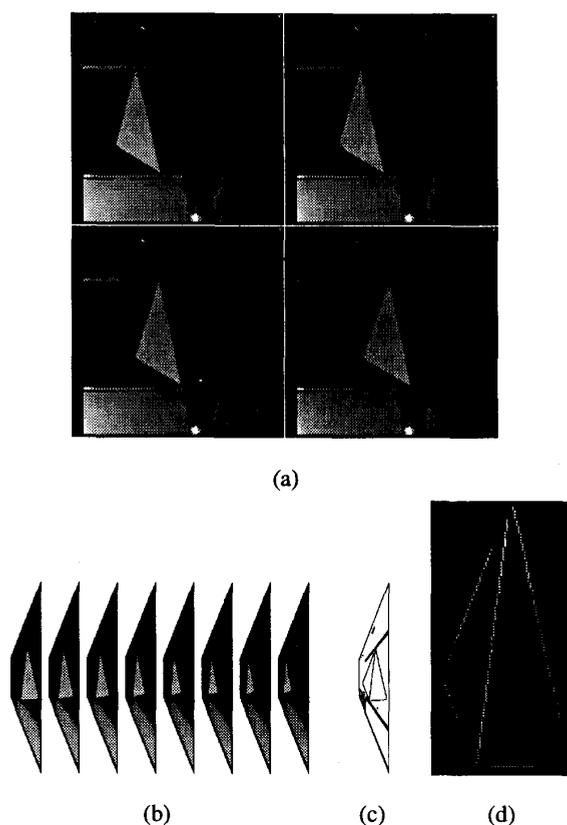


Fig. 14. (a) Ordinary lateral motion stereo images of a pyramidal block on a moving belt. Images 1, 3, 5, and 8 of the 8-snapshot sequence are shown. (b) Software-generated RWT images. (c) Edge map of the first RWT image. (d) Gray-level coded depth map for the pyramidal block from variable-resolution lateral motion stereo.

situation in the CMU image sequence which apparently had the camera pointing slightly toward the ground.

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

In the images in Fig. 13(b), some portions of the trees and buildings are not shown, because they are either above the FOE or too close to the singularity ( $x = 0$ ) line to be included. The rest of the scene is very well captured in all the RWT images. One can also observe the advantage of the variable-resolution imaging in this example as the excessive details in the near side of the road which are not quite relevant to driving are averaged out in the coarse resolution periphery of the RWT images.

The algorithm described in Sections V-A and V-C is supplemented. The correspondence ambiguities are resolved suc-

cessfully and good depth recovery results are obtained. Fig. 13(c) shows the gray-level coded depth map. In Fig. 13(d), the RWT depth map is transformed back to the uniform-resolution  $x$ - $y$  space so that the relationship to the original road image can be better apprehended. Note that the depth values of the traffic cones, the trash can and the tree trunks are correctly resolved.

## D.2 Lateral Motion Stereo

For obtaining lateral motion stereo images in our lab, a pyramidal wooden block is placed on a conveyor belt that moves from left to right. A sequence of eight snapshots (each has a size of  $512 \times 512$  pixels) from a conventional CCD camera is used in the experiment (Fig. 14(a)) since the RWT camera is not available yet. As before, the RWT images (Fig. 14(b)) are generated in software by mapping the original images from  $x$ - $y$  to  $u$ - $v$ . The middle point of the left boundary of the  $x$ - $y$  image plane is used as the origin for this mapping. The area of the resulted RWT images is approximately  $1/10$  of the original images.

Gradient-based edge detection is first performed on the RWT images. Fig. 14(c) shows the edge map from the first RWT image. Collinear points in the  $\omega$ - $\tau$  plane are detected and their  $\tau$ -intercept yields the depth, and indirectly the disparity. The voting in the  $uvd$  accumulator space results in clusters yielding the correct disparity. Fig. 14(d) displays the depth map. The result shows that the disparity changes along the edges of the pyramid are mostly computed successfully.

## VI. CONCLUSION

In this paper, a new transform, the Reciprocal-Wedge Transform (RWT), is presented as an alternative model to the log-polar transform for space-variant sensing. Exploiting the polar coordinate representation, the log-polar does well on centric rotational and scaling transformations. It, however, complicates linear features and translational transformations. Complementary to the log-polar, the RWT preserves linear features in the image, and its anisotropic variable resolution is suitable for directional space-variant sensing for problems which are translational in nature, such as stereo and linear motion. For polynomial curves, the RWT retains the form and order of the polynomial, and, hence, does not increase the complexity.

A concise matrix representation is presented. Properties of the RWT in geometric transformations are described. Moreover, the interesting insight of the projective model lends the transform to its potential hardware implementation.

As initial applications, the first simple example of road navigation shows that the perspective distortion of the road image is readily corrected by the variable resolution of the RWT which enables a more efficient search of the road direction on the reduced data. The next two examples show that the RWT is also suitable for recovering depth in both the longitudinal and lateral motion stereo.

The advantages of using multiple ( $\gg 2$ ) images in motion stereo are the improvements of precision and error rate in the depth recovery. The obvious disadvantage is the need of processing excessive amount of image data. The primary advantage

of the proposed method of *motion stereo using RWT images* is its efficiency since the variable-resolution RWT images have a significantly reduced volume of data.

In the longitudinal motion stereo, the reciprocity of the RWT has its advantage in directly eliminating the nonlinearity in the original  $x-t$  epipolar plane. In the lateral motion stereo the introduction of the RWT in both space and time domains transforms the  $x-t$  epipolar plane in ordinary motion stereo images into an  $\omega-\tau$  plane. Nevertheless, in both cases the correspondence problem in variable-resolution motion stereo is reduced to a simpler problem of extracting collinear points in the epipolar plane, which is accomplished by a voting algorithm for accumulating multiple evidence.

The variable-resolution motion stereo offers more detail and precision in depth recovery at the fovea than at the periphery of the RWT images, which seems to be natural. Its implication to active sensing appears to be direct and would be further explored.

The work of the longitudinal motion stereo is also extended to more general ego motions, especially circular movements (rotations). The RWT mapping is shown to better preserve the circular image motion as corresponding to the original vehicle motion, which indicates that the RWT is applicable to general ego motions where viewer-centered coordinates are employed. This contrasts with the benefit of handling rotation in an object-centered coordinate system using the log-polar transform.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] J. Aloimonos, I. Weiss, and A. Bandopadhyay, "Active vision," *Int'l J. Computer Vision*, vol. 1, no. 4, pp. 333-356, 1988.
- [2] R. Bajcsy, "Active perception," *Proc. IEEE*, vol. 76, no. 8, pp. 996-1,005, 1988.
- [3] D.H. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, pp. 57-86, 1991.
- [4] J.K. Tsotsos, "On the relative complexity of active vs. passive visual search," *Int'l J. Computer Vision*, vol. 7, no. 2, pp. 127-141, 1992.
- [5] M.J. Swain and M. Stricker (eds.), "Promising directions in active vision," *Int'l J. Computer Vision*, vol. 11, no. 2, pp. 109-126, 1993.
- [6] P.J. Burt, "Smart sensing within a pyramid vision machine," *Proc. IEEE*, vol. 76, no. 8, pp. 1,006-1,015, 1988.
- [7] C. Fermüller and Y. Aloimonos, "The role of fixation in visual motion analysis," *Int'l J. Computer Vision*, vol. 11, no. 2, pp. 165-186, 1993.
- [8] N. Ahuja and A.L. Abbott, "Active stereo: Integrating disparity, vergence, focus, aperture, and calibration for surface estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1007-1029, 1993.
- [9] E. Krotkov and R. Bajcsy, "Active vision for reliable ranging: cooperative focus, stereo, and vergence," *Int'l J. Computer Vision*, vol. 11, no. 2, pp. 187-203, 1993.
- [10] K.N. Kutulakos and C.R. Dyer, "Recovering shape by purposive viewpoint adjustment," *Int'l J. Computer Vision*, vol. 12, no. 2/3, pp. 113-136, 1994.
- [11] and K. Ikeuchi, "Planning multiple observations for object recognition," *Int'l J. Computer Vision*, vol. 12, nos. 2/3, pp. 137-172, 1994.
- [12] A.L. Yarbus, *Eye Movements and Vision*. New York: Plenum, 1967.
- [13] R.H.S. Carpenter, *Movements of the Eyes*. London: Pion, 1977.
- [14] E.L. Schwartz, "Computational anatomy and functional architecture of striate cortex: spatial mapping approach to perceptual coding," *Vision Research*, vol. 20, pp. 645-669, 1980.
- [15] G. Sandini and V. Tagliasco, "An anthropomorphic retina-like structure for scene analysis," *Computer Graphics and Image Processing*, no. 14, pp. 365-372, 1980.
- [16] C.F.R. Weiman and G. Chaikin, "Logarithmic spiral grids for image processing and display," *Computer Graphics and Image Processing*, vol. 11, pp. 197-226, 1979.
- [17] G. Sandini and P. Dario, "Active vision based on space-variant sensing," *Proc. 5th Int'l Symp. Robotics Research*, pp. 75-83, Tokyo, 1990.
- [18] M. Tistarelli and G. Sandini, "On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 401-410, 1993.
- [19] F. Tong and Z.N. Li, "The reciprocal-wedge transform for space-variant sensing," *Proc. Int'l Conf. Computer Vision*, pp. 330-334, 1993.
- [20] F. Tong and Z.N. Li, "Reciprocal-wedge transform in motion stereo," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,060-1,065, San Diego, 1994.
- [21] J. Van der Spiegel, et al., "A foveated retina-like sensor using CCD technology," C. Mead and M. Ismail, eds., *Analog VLSI Implementation of Neural Systems*, pp. 189-211, Kluwer, 1989.
- [22] G. Kreider and J. Van der Spiegel, et al., "The design and characterization of a space variant CCD sensor," *SPIE Vol. 1381 Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, Boston, 1990.
- [23] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Royal Society of London, Series B*, vol. 204, pp. 301-328, 1979.
- [24] R. Nevatia, "Depth measurement by motion stereo," *Computer Graphics and Image Processing*, vol. 5, pp. 203-214, 1976.
- [25] A.S. Rojer and E.L. Schwartz, "Design considerations for a space-variant visual sensor with complex-logarithmic geometry," *Proc. 10th Int'l Conf. Pattern Recognition, Vol. II*, pp. 278-285, Atlantic City, 1990.
- [26] R.O. Duda and P.E. Hart, "Use of the Hough transform to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11-15, 1972.
- [27] F. Tong and Z.N. Li, "A camera model for reciprocal-wedge transform. *Image and Vision Computing J.*, (submitted), 1994.
- [28] C. Thorpe, M.H. Hebert, T. Kanade, and S.A. Shafer, "Vision and navigation for the Carnegie-Mellon Navlab," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 3, pp. 362-373, 1988.
- [29] E.D. Dickmanns and B.D. Mysliwetz, "Recursive 3D road and relative ego-state recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 199-213, 1992.
- [30] R.A. Lotufo, A.D. Morgan, E.L. Dagless, D.J. Milford, J.F. Morrissey, and B.T. Thomas, "Real-time road edge following for mobile robot," *Electronics & Communications Engineering J.*, vol. 2, no. 1, pp. 35-40, 1990.
- [31] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353-363, 1993.
- [32] R.C. Bolles, H.H. Baker, and D.H. Marimont, "Epipolar-plane image analysis: an approach to determining structure from motion," *Int'l J. Computer Vision*, vol. 1, pp. 7-55, 1987.

**Frank Tong** received the BS degree in computer science from the Chinese University of Hong Kong in 1983, and the MS degree in computing science from Simon Fraser University, Canada, in 1987. He is currently completing a PhD degree in computing science at Simon Fraser University. Since August 1994, he has also been a member of the Computer Information Systems faculty at University College of the Fraser Valley. His research interests include active vision, space-variant sensing, parallel vision algorithms, pattern recognition, and robotic applications.

**Ze-Nian Li** received the BS degree in electrical engineering from the University of Science and Technology of China in 1970, and the MS and PhD degrees in computer sciences from the University of Wisconsin at Madison in 1981 and 1986, respectively.

From 1970 - 1979 he was an electronic engineer in charge of digital and analogical system design. He was an assistant professor at the University of Wisconsin at Milwaukee from 1986 - 1987. In 1988 he joined the School of Computing Science at Simon Fraser University, where he is currently an associate professor. His research interests include computer vision, pattern recognition, and parallel machines and algorithms for image processing and analysis.

