Reciprocal-Wedge Transform in Active Stereo *

Ze-Nian Li and Frank Tong
School of Computing Science
Simon Fraser University
Burnaby, B.C.
CANADA V5A 1S6

Phone: (604) 291-3761
Fax: (604) 291-3045
e-mail: li@cs.sfu.ca

Abstract

The Reciprocal-Wedge Transform (RWT) facilitates space-variant image representation. In this paper a V-plane projection method is presented as a model for imaging using the RWT. It is then shown that space-variant sensing with this new RWT imaging model is suitable for fixation control in active stereo that exhibits vergence and versional eye movements and scanpath behaviors. A computational interpretation of stereo fusion in relation to disparity limit in space-variant imagery leads to the development of a computational model for binocular fixation. The vergence-version movement sequence is implemented as an effective fixation mechanism using the RWT imaging. A fixation system is presented to show the various modules of camera control, vergence and version.

Keywords: Active vision, eye movements, Reciprocal-Wedge Transform, space-variant sensing, stereo

*This work was supported in part by the Canadian National Science and Engineering Research Council under the grant OGP-36726.

1 Introduction

Stereopsis is one of the most studied areas in computer vision. Computer algorithms computing the stereoscopic disparity can be dated back to Marr and Poggio's work of applying various constraints to the so called *correspondence* problem [1]. Notwithstanding the persistent efforts of many fine researchers, the stereo correspondence problem still remains one of the difficult problems to be solved. The difficulty is often due to the ambitious goal of a reconstruction of complete depth maps from limited static views. As pointed out by Fermüller and Aloimonos [2], it has become clear that active (and selective) acquisition of data is essential to stereo vision. Moreover, stereo correspondence is often linked to the fusion of two disparate uniform resolution images. As the methods devised are mostly for accurate recovery of the image disparity, the process can be considered as computing the foveal fusion in the domain of space-variant sensing. However, the structure and functional objective of the peripheral vision are different from those for foveal processing. The issues of peripheral fusion have not received much attention. This is in part due to the lack of research in stereo vision using anthropomorphic sensors.

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

In human vision, the binocular fixation is accomplished by two components — *version* and *vergence* [3].

Version is the conjugate movement of the eyes. Version movements are similar in amplitude and direction in the two eyes, and thus obey Hering's principle of "equal innervation" [4]. Pure version occurs when the gaze is transferred under zero disparity from one object to another. It requires that the two eyes maintain their convergence while panning synchronously at the same angle in the same direction. Version is the fast saccadic movement of the two eyes. In fact, the movement is so fast that there is no time for visual feedback to guide the eye to its final position. Sometimes, the magnitude of the velocities can reach more than

700° per second for large amplitudes [3].

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

Experiments have shown that the human stereopsis accepts only a very limited range of disparities. The Panum's area forms a limited zone about the fixation point. Beyond the Panum's area, human vision systems can no longer fuse the stereo images. Burt and Julesz [5] conducted some experiments on fusion in the context of disparity gradient. They made an amendment to the previous understanding of Panum's fusional area, i.e, binocular fusion occurs only when the disparity gradient does not exceed a critical value of ~ 1.

Stereo problems are greatly simplified in verging systems because vergence control allows redistribution of the scene disparities around the fixation point, thus reducing the disparities over an object of interest to near zero. Olson [6] presented a simple and fast stereo system that is suitable for the attentive processing of a fixated object. In view of the narrow limits of the Panum's area, the fusible range is thought to be a privileged computational resource that provides good spatial information about the fixation point. Assuming vergence control, Olson's stereo algorithm capitalizes on a restricted disparity range.

Pahlavan, Uhlin and Eklundh [7] developed their machine fixation model and the KTH head-eye system after the fixational behaviors in human vision. Two types of vergence movements are studied, i.e., accommodative vergence and disparity vergence. They are supported by the integration of the blur and disparity stimuli.

Grosso and Tistarelli [8] reported an active and dynamic vision system that combines binocular stereo and motion. Their active control strategy features camera fixation and tracking of a point in space, and therefore reduces the need for camera calibration under robot motion and head rotation.

Other works combine different cues to perform active camera control for stereo ranging.

Krotkov and Bajcsy [9] developed and implemented the idea of cooperative ranging in their agile stereo camera system. Krotkov's system demonstrates the reliability in ranging upon fusion of the focusing and stereo vergence components. Abbott and Ahuja [10] took integration of visual cues to great length in their University of Illinois Active Vision System. Complementary strengths of different cues are exploited in integration via active control of camera focus and orientation, as well as aperture and zoom settings, thus coupling image acquisition and surface estimation dynamically and cooperatively in an active system.

Among the few results reported for stereo disparity in the domain of space-variant sensing, Griswold, Lee and Weiman [11] described a fusion model using the log-polar imaging. Disparity values were computed by convolving the binary edge images with filter templates established for different edge lengths, disparity widths and edge orientations. The inevitable image distortions after the log-polar transform caused hardship in their method.

The *Reciprocal-Wedge Transform (RWT)* was proposed [12] for spatially variable-resolution sensing in support of active vision. The RWT exhibits nice properties in computing geometric transformations owing to its concise matrix notation. It supports variable-resolution sensing and thus facilitates efficient data reduction. Unlike the log-polar transform, 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.

The RWT has initially been applied to road navigation and motion analysis [12, 13] to illustrate the effectiveness of the new imaging model. The longitudinal and lateral motion stereo algorithms benefit from the perspective correction, linear feature preservation and efficient data reduction of the RWT. The longitudinal motion stereo algorithm is later extended [14] to work under circular ego motions performed by a NOMAD 200 mobile robot [1] in a laboratory environment. Recently, it is shown [16] that a hardware RWT camera can be designed to capture the RWT images instantly.

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

[1]NOMAD 200 is manufactured by the Nomadic Technologies Inc., California, USA. Additional CCD camera and pan-tilt camera platform are mounted for our active vision research.

the RWT. A scanpath experiment, inspired by the eye movement research, demonstrates a desirable performance of our fixation system. It is shown that the RWT imaging model is suitable for fixation control in active stereo. The RWT images fulfill the requirement of foveal-peripheral variable resolution. They also facilitate the computation of stereo disparities which are manifested as lateral image translations. Compared to the traditional reconstructionist approach, active exploration visual behavior is shown to be beneficial.

The organization of the rest of the paper is as follows. Section 2 reviews the RWT and presents a new V-plane projection method for RWT imaging. Section 3 discusses the two types of eye movements in binocular vision and our RWT computational model for binocular fixation. Section 4 describes our experiments and the results. Section 5 concludes the paper.

2 Reciprocal-Wedge Transform

The Reciprocal-Wedge Transform (RWT) was proposed as a new imaging representation for space-variant sensing [12]. The RWT maps a rectangular image into a wedge-shaped image. Mathematically, the RWT is defined as a mapping of the image pixels from the x-y space to a new u-v space such that

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

The lady and the grid images in Figure 1 illustrate the forward and backward RWT. Note the blurring at the periphery of Figure 1(c) and 1(f), which results from the significant data reduction after the RWT (sometimes $> 95\%$ [12]). In Figure 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.

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

$$\mathbf{T} \;=\; \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \;=\; \mathbf{T}^{-1} \;, \tag{2}$$

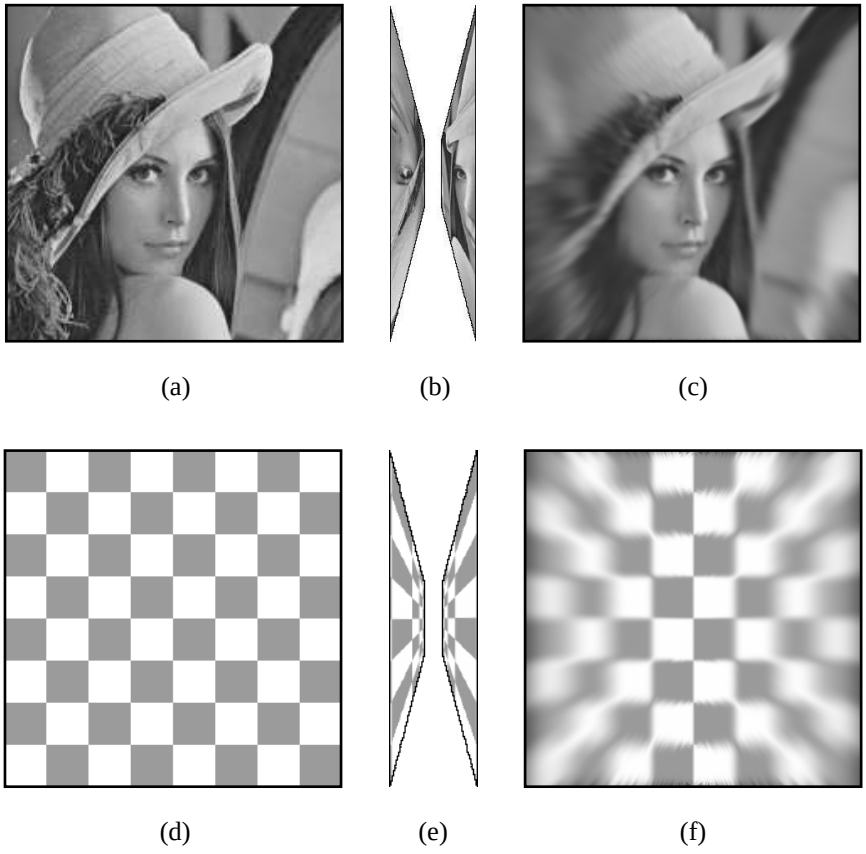

Figure 1: The Reciprocal-Wedge transform. (a) The lady's image. (b) The RWT image shows two inside-out wedges. (c) The image when transformed back to the Cartesian domain. (d) A rectangular grid. (e) The RWT image. (f) The grid transformed back to illustrate the resolution varying from the center to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

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

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

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

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

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

2.1 Shifted Reciprocal-Wedge Transform

The singularity of the RWT exists at $x = 0$, i.e., $u = 1/0 = \infty$ and $v = y/0$. The *shifting* method is one way of fixing the singularity problem. It is to introduce a shift parameter a in the RWT.[2] This variant formulation is called *Shifted Reciprocal-Wedge Transform (S-RWT)* [12].

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

Both the forward and backward transformations for the S-RWT remain the same cross-diagonal matrix (eq. (2)) except the additional parameter a.

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

The effect of the parameter a is to horizontally shift the center strip (and the rest of the image) away from $x = 0$, or equivalently, shift the x axis in the Cartesian image. (The RWT images in Figure 1 are indeed generated in this way.) The parameter a should be of opposite sign for the left and right halves of the Cartesian image, i.e., the two halves of the image are respectively shifted in opposite directions.

[2] A similar shift parameter is also used in log-polar transform to the same effect [15].

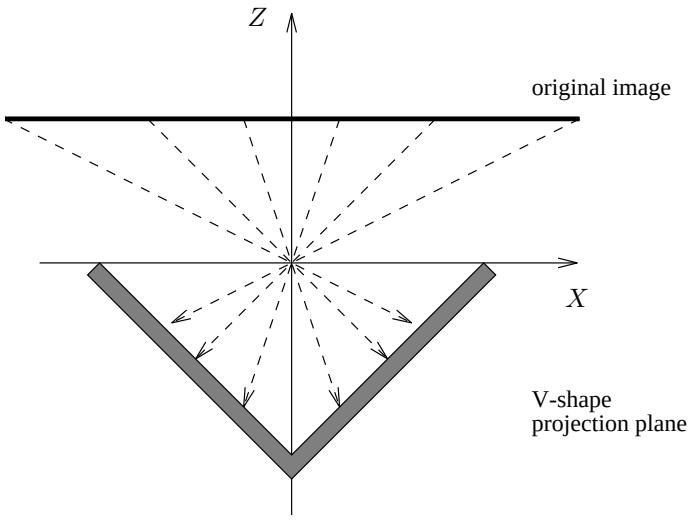

Figure 2: V-plane projection. The left arm of the V forms the projection plane for the right half of the original image and the right arm of the V is for the left half. The singularity problem is resolved, and space-variant resolution is effected on both projection planes.

2.2 V-plane Projection for S-RWT

The RWT can be realized by a projection onto non-frontal imaging planes [12, 16]. However, it requires a delicate optic design as shown in [16]. This section shows that the S-RWT can be more readily implemented with a V-plane projection (Figure 2).

The two halves of the projection plane are joined to form a V in this figure. The left arm of the V forms the projection plane for the right half of the original image and the right arm of the V is the projection plane for the left half. The singularity problem disappears because the center region of the original image gets projected to a u position on the V-plane.

A point P on the original image is projected to Q on the projection plane. O is the center of projection, and E is the origin of the x-y space. To be consistent with the S-RWT formulation in eq. (3), the origin of the u-v space F is defined as the point of projection when $x = \infty$ and $y = 0$.

From the geometry in Figure 3,

$$\frac{\overline{RE} + x}{\overline{RF} + u} = \frac{\overline{OF}}{u} \quad . \tag{4}$$

Since $\overline{RE} = r\cos\theta$ and $\overline{OF} = \overline{RF} = \frac{r}{2\cos\theta}$,

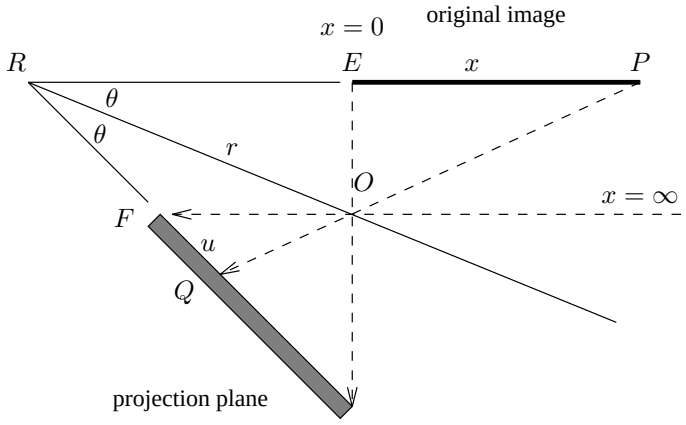

Figure 3: Geometry of the V-projection from P to Q.

$$u \;=\; \frac{(\frac{r}{2\cos\theta})^2}{x + (r\cos\theta - \frac{r}{2\cos\theta})} \;=\; \frac{f^2}{x+a} \tag{5}$$

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

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

$$\frac{\overline{PR}}{\overline{OF}} \;=\; \frac{x + r\cos\theta}{\frac{r}{2\cos\theta}} \;, \tag{6}$$

$$\frac{\overline{PQ}}{\overline{OQ}} \;=\; \frac{\overline{PO}}{\overline{OQ}} + 1 \;=\; \frac{y}{v} + 1 \;. \tag{7}$$

Combining (6) and (7),

$$v \;=\; \frac{\frac{r}{2\cos\theta} y}{x + (r\cos\theta - \frac{r}{2\cos\theta})} \;=\; \frac{fy}{x+a} \;. \tag{8}$$

From (5) and (8), it shows that the u and v coordinates from the V-plane projection are effectively computing the S-RWT as defined in eq. (3) within a constant factor f.

3 Binocular Vision in Space-variant Sensing

In this section, we shall investigate the Panum's fusion in the context of space-variant binocular sensing. Specifically, the computational view of the Panum's fusional area in the RWT sensing space based on the V-plane projection imaging method will be studied, and the fixation mechanism in the RWT binocular vision will be presented.

3.1 Fusional Range in RWT

Objects on the horopter form stereo images on the corresponding retinal elements in the two eyes. Images of zero disparity as such are perfectly fusible, and are seen single. Panum (1861) showed that zero disparity is not the necessary condition for singleness [17]. An image on one eye would fuse with a similar image on the retina of the other eye within a small area about the corresponding point. The interval for the object position where no doubling is seen defines the limits of the Panum's fusional area.

Quantitative studies by Fischer (1924) and Ames (1932) yield data that plot out the size of the Panum's area at different visual angles [17]. Fender and Julesz [18] reported that binocular fusion occurs in regions vary from 6 min. of arc at the center of the visual field to 20 min. of arc at the peripheral angle of 6°. The rapid dilation of the Panum's area at the peripheral visual angles can be functional in nature. When one interacts with the environment, accurate foveal processing serves well for attentive inspection of the fixated target. However, general monitoring of the wide visual field is obviously important for detection of activities, smooth maneuvering and the spatial percept of the external environment.

In computer vision, the fusional range is computationally modeled by disparity limits. We address the issue of variable Panum's area in relation to the RWT space-variant resolution. In the following, a binocular system of RWT cameras is studied. We set up the projection equations and fed them to Maple V [19] (a numerical software for scientific computation) to obtain the plots of the disparity contours for the different fusional limits.

Figure 4 gives a schematic diagram of the RWT binocular system. The cameras are placed symmetrically on the two sides about the Z-axis, with their nodal points on the X-axis, and imaging the positive Z half-space. Let $2b$ be the baseline separation of the cameras. The focal length of the cameras is denoted by f, and the inter-projection-plane angle is 2ψ. The cameras are fixating the point Z_o on the Z-axis. Let P be a point located at (X, Y, Z); u_l and u_r are the RWT coordinates of the left and right images of P respectively. Let the disparity be denoted by d. The triangulation geometry in Figure 4 yields the following equations:

$$\frac{-u_l}{\sin(\theta_l - \phi)} = \frac{f}{\sin(\theta_l - \phi + \psi)} ,$$

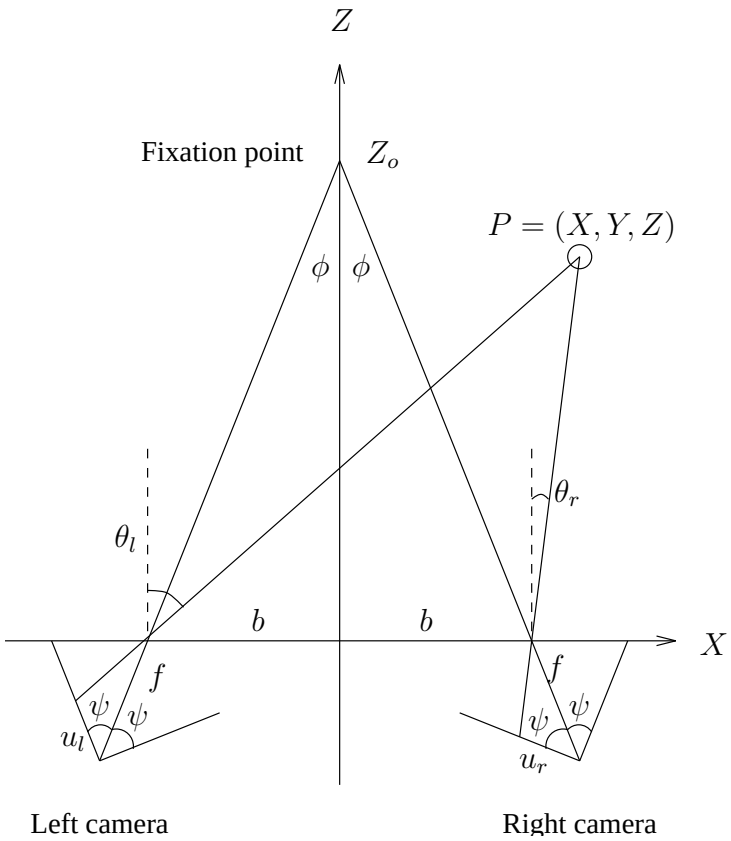

Figure 4: An RWT binocular system.

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

The system of equations are solved for X and Z at different disparity values, d. For a typical imaging situation, set $b = 200$, $f = 200$, $\phi = 45°$, and $Z_o = 6000$ (in 1/100 inch unit). The numerical values of X and Z are calculated for d ranging from -4 to $+4$. Figure 5 plots the (X, Z) coordinates for $d = 0, \pm 2, \pm 4$. Each of the curves represents a disparity contour of a particular d. All points on the same contour will form disparate images in the two RWT cameras with the disparity d. These contour are due to the specific imaging configuration of the RWT binocular system. However, the corresponding fusional region indeed exhibits the

desired property of fovea-periphery variable extent. In this example, the fusional region at the peripheral angle of 36° is twice as deep as that at the central position.

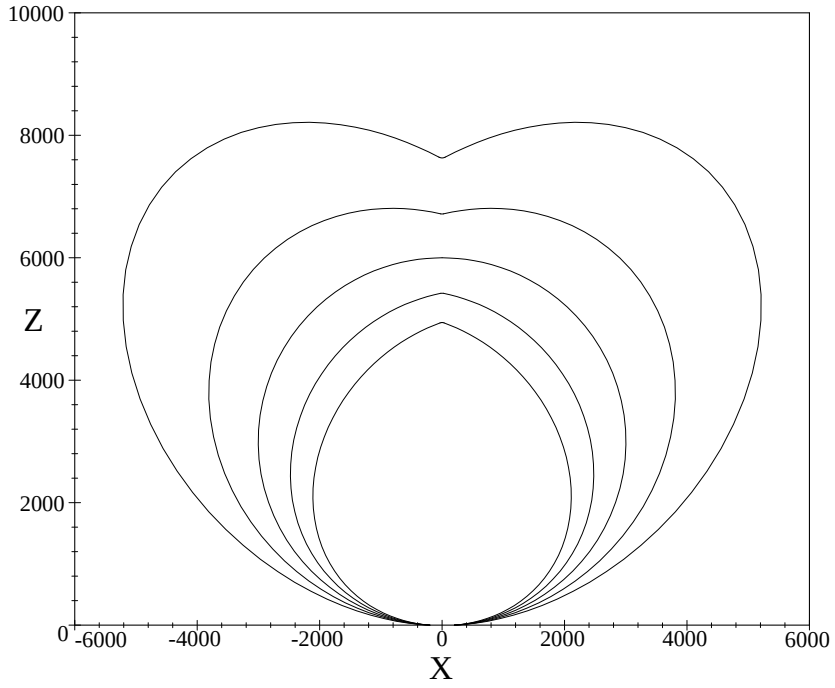

Figure 5: Disparity contours for the RWT binocular projection. The plot is obtained by setting the baseline separation $2b = 400$, the focal length $f = 200$, and the inter-plane angle $\phi = 45°$. The cameras are converged at the fixation point of 6000 on the cyclopean axis. From the outermost contour to the innermost one, the disparity contours are plotted in the order of $d = +4, +2, 0, -2, -4$.

Comparison of the RWT binocular system are drawn with the conventional uniform-resolution cameras. The model of a verging system of uniform-resolution cameras is given in Figure 6. This time, the set of equations yielded read as follows:

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

Again, the system of equations are solved in Maple V for X and Z. Similarly, a plot of the (X, Z) coordinates is performed for $d = 0, \pm 2, \pm 4$, with the settings of $b = 200$, $f = 200$, and $Z_o = 6000$ (Figure 7).

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

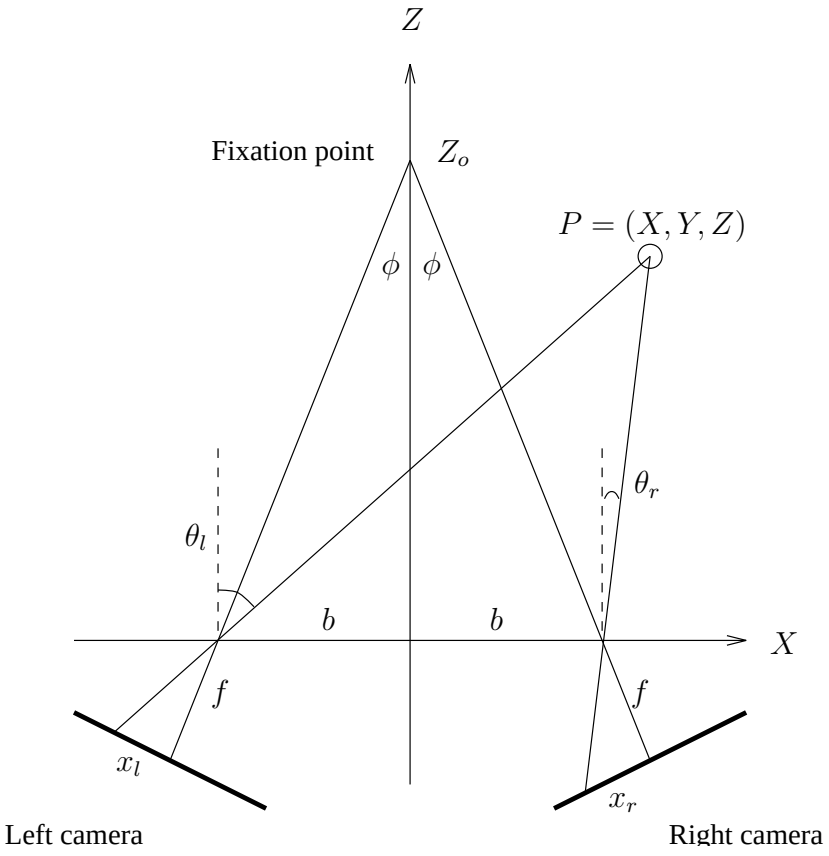

Figure 6: A verging system with uniform-resolution cameras.

3.2 Fixation Mechanism

In this section, the vergence and version camera movements are computed using space-variant image representation. Human retina is a space-variant sensor of a fovea-periphery structure. Experiments show that when one changes fixation to a nearer target point, the

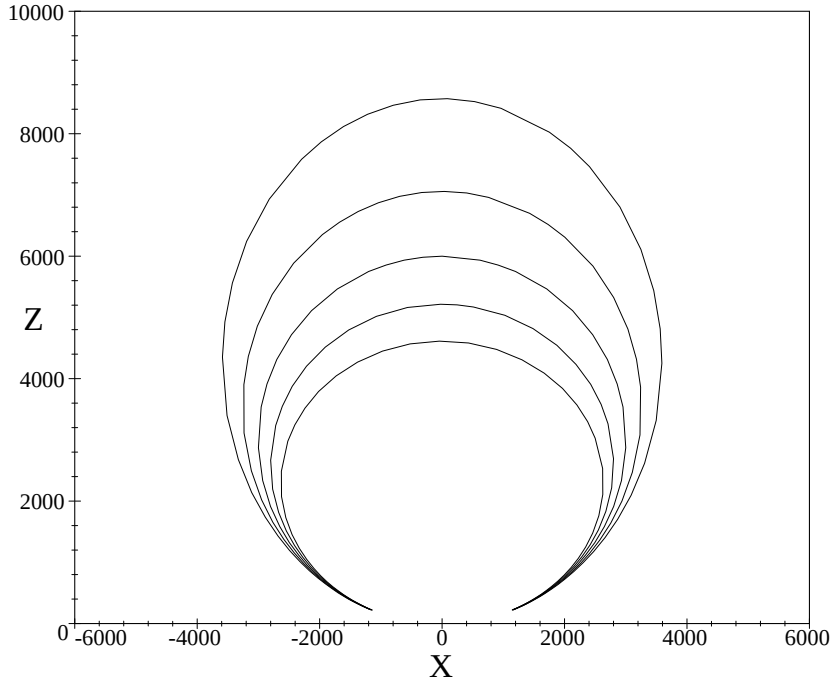

Figure 7: Disparity contours for uniform-resolution cameras. The plot is obtained by setting the baseline separation $2b = 400$, the focal length $f = 200$, and the fixation distance $Z_o = 6000$. From the outermost contour to the innermost one, the disparity contours are plotted in the order of $d = +4, +2, 0, -2, -4$.

two eyes first undergo a symmetrical vergence to bring the fixation nearer to the target. In the middle of the vergence movement, a conjunctive saccade is superimposed to swing the gaze in line with the target. The vergence then proceeds to completion in the final stage to bring the fixation accurately to the target [20]. We argue that from the computational point of view, the three-phase vergence-version fixation movement is an efficient mechanism with the space-variant sensor.

Consider the case when the cameras are fixating an object A in the scene, and is about to change gaze to a nearer object B at periphery. A is fixated in the fusional region at the fovea. B, although located in the periphery, is covered in a deeper fusional area. Computationally, the fusional area's limit is used to the advantage for restricting the disparity range. Under the limited operating range for disparity, B's disparity is readily resolvable even though its depth differs very much from that of the fixation. If the cameras were straightforwardly

gazed at B at this time, B might become out of the fusional limit when it is brought into the foveal direction. The depth of B would be difficult to calculate and the fixation would fail. A more effective mechanism is to have a first vergence to change the fixation distance so that B is lying close to the horopter after the vergence. This also prepares for the versional movement so that when B is brought to the foveal direction, it will still be imaged within the fusional limit. Next, based on the rough estimate of B's visual angle, a pan movement is launched to direct both cameras to the direction near B. Now, B is in a near-foveal direction, and located within the fusional limit. This is true owing to the first vergence. Finally, a second vergence can be executed to bring B accurately into fixation. Figure 8 summarizes the camera movement of a space-variant binocular sensor.

3.3 Disparity Computation in RWT

Another property that renders RWT suitable for stereo vision is the anisotropy of its space-variant resolution. In stereo vision, the disparate images formed in the binocular cameras differ from each other by a horizontal displacement. In the conventional images, disparity is computed by correlation along the horizontal dimension. A rectangular pattern in the Cartesian image appears as shifted along the horizontal streamlines (Figure 9(a)). The RWT maps the horizontal streamlines into radials in the RWT domain. Figure 9(c) shows the bipolar RWT image. The radial streamlines converge at the two antipodes on the u-axis. In the RWT image, the rectangular pattern is transformed into a wedged rectangle displaced along the radial streamlines.

Disparity computation may become very complicated in other schemes of image representation. In the log-polar model, horizontal streamlines are mapped to complicated log-sine curves (Figure 9(b)). The difficulty is at least two-fold. First of all, disparate images are not related in a linear structure any more. Search for stereo correspondence has to be conducted along these log-sine curves which are expensive to compute. In addition, the image pattern gets rotated and scaled while being translated along the log-sine curve. A complicated procedure is required to calculate the image motion in order to make it possible for a correlation operator to be used for the disparity computation [11].

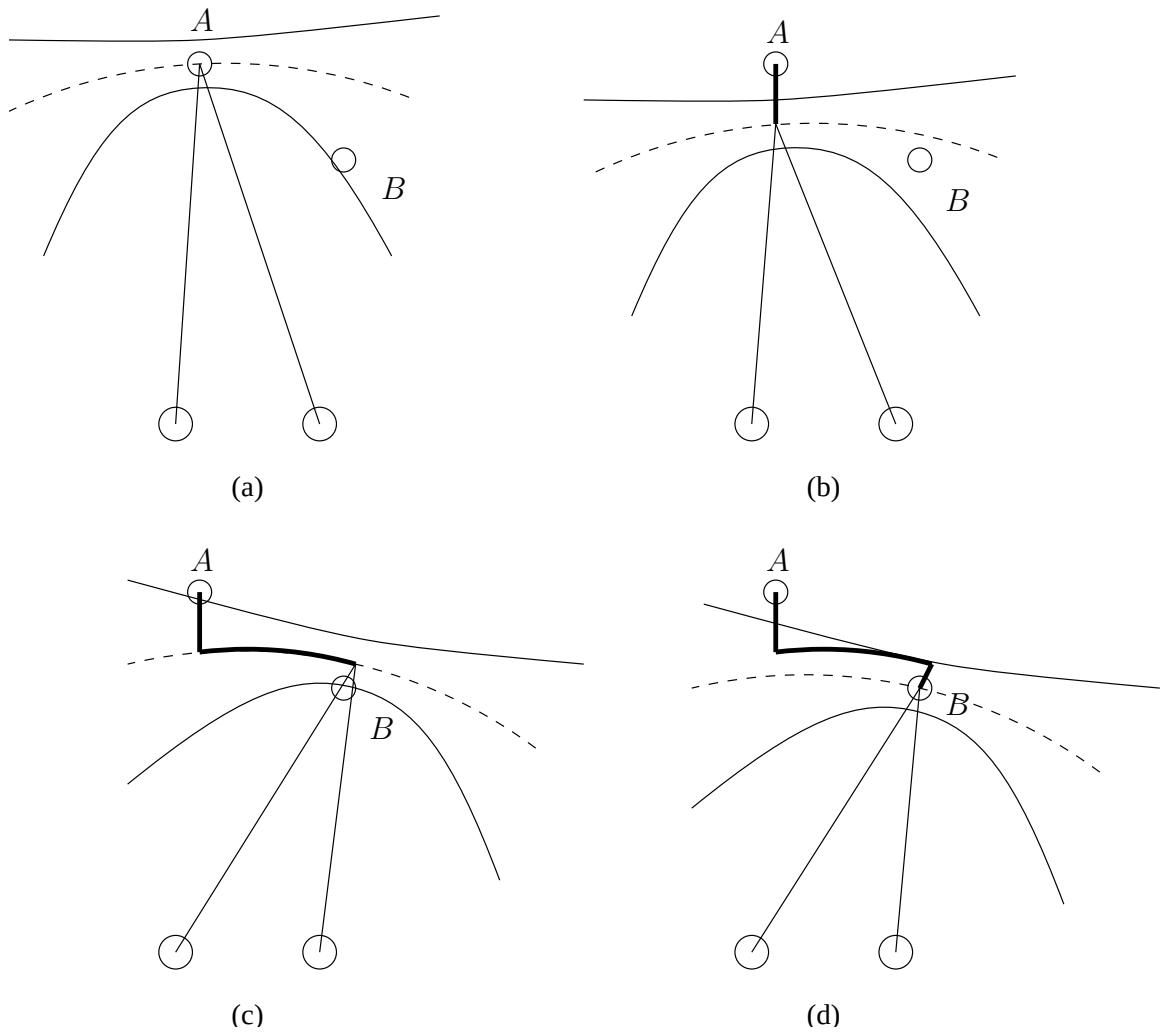

Figure 8: Ocular movement of space-variant binocular sensor. (a) The cameras are fixating A. (b) First vergence brings the fixation point to close to B's depth. (c) Version brings the cameras in line with B. (d) Second vergence, the cameras fixate precisely on B.

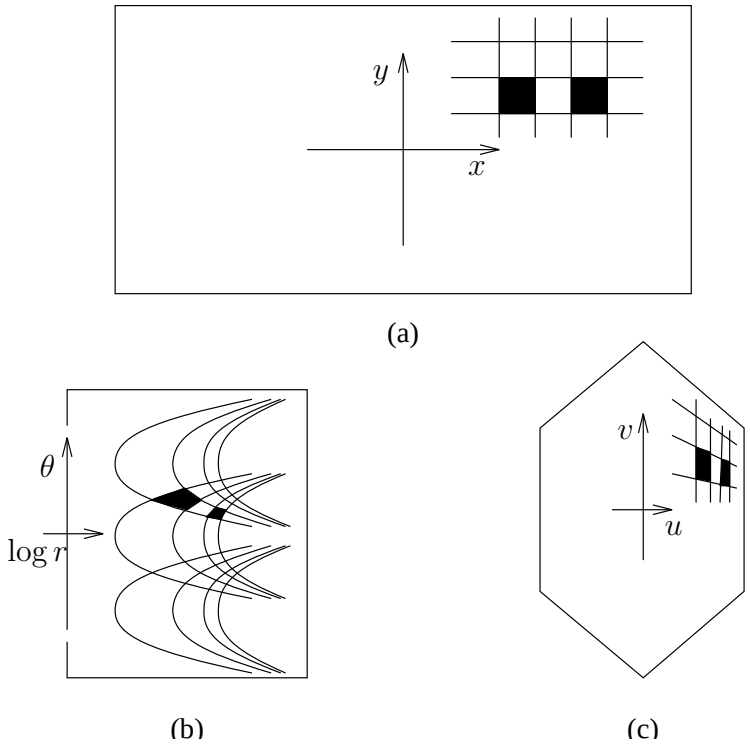

Figure 9: Disparity in different image representations. (a) Disparity is manifested in horizontal translation in the Cartesian image. (b) Horizontal translation becomes a complicated image motion in the log-polar domain. (c) Horizontal translation is mapped to translation along the radial streamlines in the RWT image.

The mapping of corresponding points into RWT domain does not introduce rotations, which facilitates an easier search of matching points in the left and right images. If d is the image disparity, the left and right RWT image coordinates are related as follows:

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

3.4 A Binocular Fixation System

This section describes the design of a system (Figure 10) for the interactive fixation process described above. It comprises the vergence and version components interfacing with the controller of the camera pan-tilt platform. The next fixation which initiates vergence and version oculomotor sequence is computed by the "where-next" component. Vergence is a slow and visually guided process. It is adjusted according to the disparity, thus completing

the feedback loop.

The camera platform houses two cameras each of which has the two degrees of freedom for pan and tilt respectively. In the design, the cameras are RWT cameras which output RWT images of the scene directly. For now, ordinary cameras are used and the RWT images are generated from the uniform-resolution images with a Reciprocal-Wedge transformation routine. The gaze angles for vergence and version are mapped to the mechanical movements of pan and tilt for individual camera. The version angle drives identical movements of pan and tilt for both cameras, whereas the vergence is split evenly into disjunctive convergence or divergence between the two cameras.

The component "where-next" represents the high-level intelligent process for selecting the next fixation point in the scene. The left and right RWT images are combined to yield a cyclopean image of the scene. The "where-next" component searches in this cyclopean image for features of interest. In fact, the next-fixation computation is a highly involved process [20]. Although this high-level intelligent process for computing the next fixation is an interesting topic for research, it is beyond the scope of this paper. In the following demonstration of an active fixation system, simplistic heuristic criteria are used to show the usual scanpath behavior in binocular visual exploration.

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

In Figure 10, two disparity modules are simulated, namely the peripheral disparity and foveal disparity described above. The former is used in the first vergence to eliminate the peripheral disparity. The latter is used in the second vergence to converge precisely on the

target inside the fovea.

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

4 Experimental Results

An office scene (Figure 11) is used for experiments of fixation transfer and scanpath demonstration. Stereo images of the scene are captured in uniform resolution (512×512) using ordinary cameras. When converted and resampled to RWT images, the amount of data is reduced by approximately 80%.

4.1 Fixation Transfer

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

In the RWT binocular system, when changing from the current fixation to another target at the visual periphery, the model for camera movement described in Section 3.2 is followed. Upon changing gaze from the current fixation point to the next target, the target may be located well within the fusional limit at the periphery under the variable fusional area. Thus, a rough estimate for the target's peripheral disparity can be calculated. The two cameras are then converged/diverged to reduce this disparity. This corresponds to the first vergence movement. Next, the cameras are panned to the viewing angle of the target to bring the target to the fovea for higher resolution imaging. This corresponds to the versional movement. Now, the target is in the foveal direction of the cameras but it is likely imaged with a residual foveal disparity. In this phase of the fixational movement stereo matching is performed in the fovea to obtain the foveal disparity and the cameras are converged/diverged

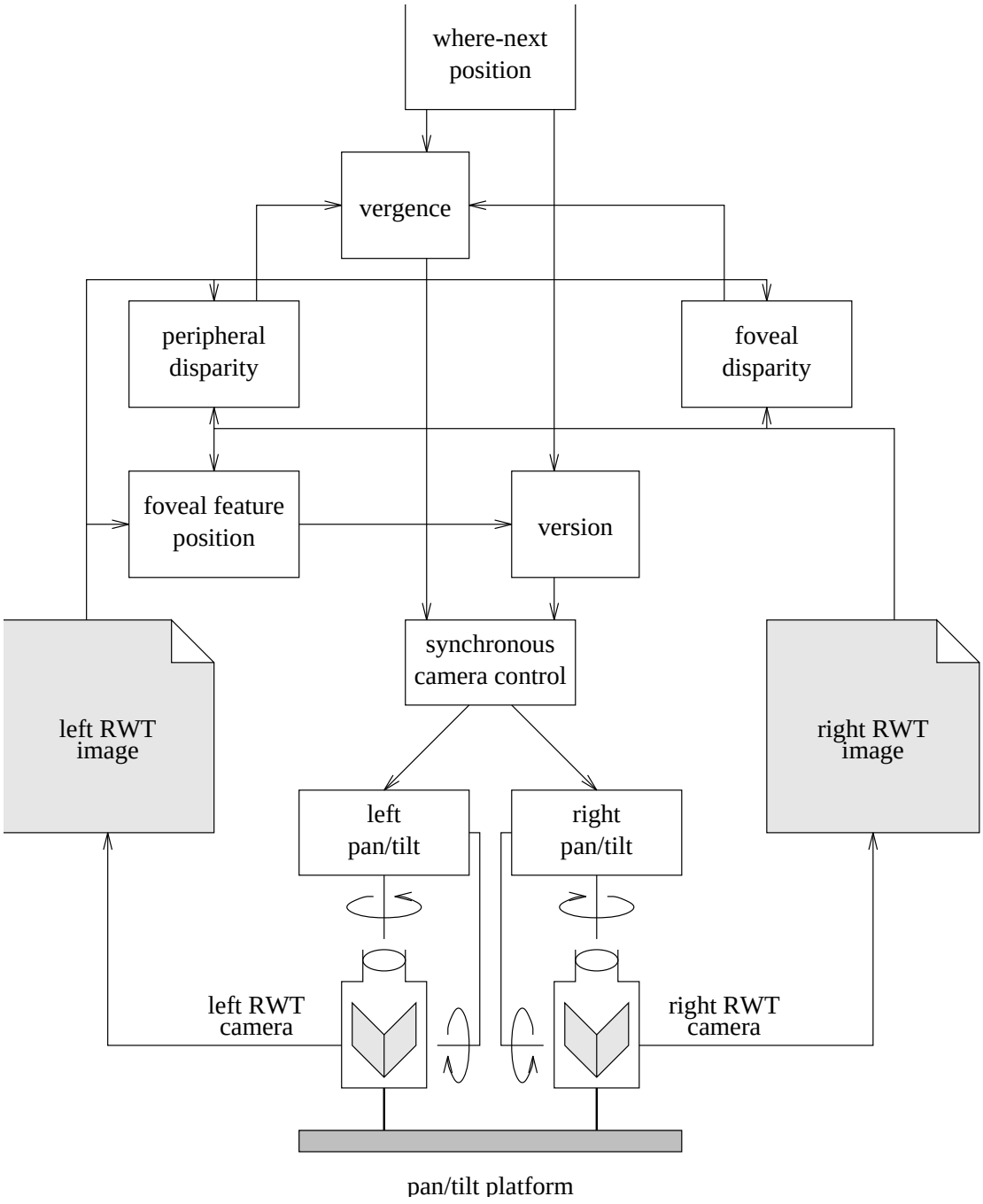

Figure 10: An interactive fixation system.

Figure 11: An office scene.

to zero in on the target precisely. This movement corresponds to the second vergence.

Figure 12(a-d) shows a test of the fixation process in the RWT binocular system. Figure 12(a) shows the images corresponding to a fixation on the computer keyboard in the office scene. It shows the RWT images of the scene and the disparity map. The left and right edge maps in Cartesian coordinates are superimposed and shown here for the reader's apprehension of the disparate scene images and the camera orientations.

As the chair is located at a closer range to the cameras in relation to the keyboard (the current fixation point), it exhibits a non-zero disparity. The disparity value, however, is small as it is located in the periphery. The image disparities in this example are well within the fusional limit. The disparities are computed by applying a 7×7 windowed correlation over a range of $[-5, 5]$. The disparity results reveal different disparities for objects at different depth from the keyboard. The chair has a large crossed disparity whereas the magazine organizer on the desk shows a non-zero uncrossed disparity[3].

In this experiment, preference for the next gaze is given to the one with larger disparity, hence the gaze is changed from the computer keyboard to the chair. Three intermediate steps are involved. Initially, the cameras are fixated at the keyboard. A disparity of -4 is detected with the chair at a peripheral angle corresponding to $u = -72$ pixels. To simulate

[3]We have chosen to show the absolute disparity values in all grey-level coded disparity maps. In other words, both disparities -5 and 5 are shown to be the brightest while 0 is shown dim on a completely dark background. The advantage is so that the difference between zero disparity and non-zero disparities is clearly displayed. The disadvantage is that crossed and uncrossed disparities can not be distinguished

Superimposed Edge Maps

left RWT right RWT disparity map

Figure 12: (a) Fixation sequence. Initially, fixation is on the computer keyboard.

Superimposed Edge Maps

left RWT right RWT disparity map

Figure 12: (b) First vergence. the peripheral disparity of the chair becomes zero.

Superimposed Edge Maps

left RWT right RWT disparity map

Figure 12: (c) Version. The chair is brought to the fovea.

Superimposed Edge Maps

left RWT right RWT disparity map

Figure 12: (d) Second vergence. Fixation is precisely on the chair.

the camera convergence in the next step, the -4 disparity at $u = -72$ is translated back to the Cartesian domain by the inverse RWT transformation to a disparity of -10 pixels at $x = -101$. A mapping function maps the -10 disparity to the disjunctive vergence angle that converges the cameras so that the peripheral disparity of the chair image becomes zero. The RWT images are then obtained from the Cartesian scene images for the new camera orientations as though they are from the real RWT cameras. Figure 12(b) now shows the result of the first vergence. The chair images at $u = -70$ are now well aligned as seen in the edge map in (b), and the disparities shown in the disparity map demonstrate that zero disparity is achieved with the chair images.

Next, the cameras are panned to the left for an angle corresponding to 72 pixels in the RWT domain. Figure 12(c) shows the result of this conjunctive versional movement. The chair images now come to the foveal region of the cameras. It is observable that the estimate for the peripheral disparity during the first vergence is not accurate enough for high resolution processing inside the fovea. The disparity map in (c) shows that the residual disparity in the chair images becomes apparent once they are placed in the fovea. This foveal disparity, however, has a value well within the operational range of the fusional limit since the first vergence has already achieved a good approximation.

Figure 12(d) now takes the vergence to completion. The foveal disparity of the chair is computed. It is a small residual disparity of 1 pixel. The cameras are then diverged by an angle corresponding to 1 pixel in the RWT images. The disparity map in (d) shows that the cameras are precisely fixating the chair in the fovea.

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

4.2 A Scanpath Demonstration

Scanpath [21] is the sequence of fixation that one exercises during a visual scan. The scanpath behavior of the system is demonstrated in an experiment of binocular visual exploration.

Although the cognitive modeling of scanpaths is a rigorous research topic in psychology [20, 21, 22], we did not delve into the issues raised therein. Instead, at each stop simplistic heuristics are employed to determine the next fixation. The resulting scanpath is to demonstrate an operation of our fixation system.

The experiment is conducted with the image data of the office scene in Figure 12. Initially, the fixation is set on the computer keyboard on the desk. The next point of interest is chosen based on three considerations. (1) It is a sizable object worth exploring. (2) It has the most disparate image in the current scene. (This drives the system to sweep the entire depth of the scene efficiently.) (3) It has not been explored in detail as yet so that the system would not come to the same object repeatedly. The heuristics are simple enough, yet work successfully in transferring the initial fixation from the computer keyboard to the magazines standing next to the monitor. As shown in Figure 13(a), the gaze is then changed to the chair, the computer terminal, and then to the roller wheels of the chair.

The prime observation we emphasize from the outcome of this experiment is the successful working of the fixation system as a whole in implementing the fixation transfer mechanism at each fixation. For example, the initial fixation is on the computer keyboard (Figure 13(a-1)). The RWT disparity image in Figure 13(b-1) shows an extended area (325 pixels) of 2-pixel disparity occur at the position of $u = 51$ and $v = 33$ (corresponding to the magazines in the scene). The execution log of the simulation program indeed has recorded the following inter-component interactions that happened in the system.

As the "where-next" component evaluated the next fixation to $(51, 33)$, the fixation transfer routine was initiated in the vergence and version components. The first vergence was effected by a vergence control to the camera for a divergence angle corresponding to a 2-pixel peripheral disparity at the position $(51, 33)$. Then the version component was initiated with a pan-tilt corresponding to 51 right and 33 up in the RWT coordinates (equivalent to 55 right and 46 up in the Cartesian coordinates). A foveal disparity then was evaluated to -1 pixel, causing the vergence component to launch the second vergence for a convergence angle corresponding to a 1-pixel foveal disparity. Finally, an edge feature was detected by the foveal feature component at a position 2 pixels to the left of the center. This resulted

in a versional adjustment of 2 pixels, placing the fovea precisely on the edge feature (i.e., on the magazines). The result can be appreciated in Figure 13(a-2) which shows the dark edge of the magazines positioned right at the center of both stereo images. The process then continued with the "where-next" selecting position $(-90, 54)$ for the new fixation, and the fixation routine was repeated. Overall, the log records indicate the successful execution by the fixation system as a whole with correct interactions between the various components.

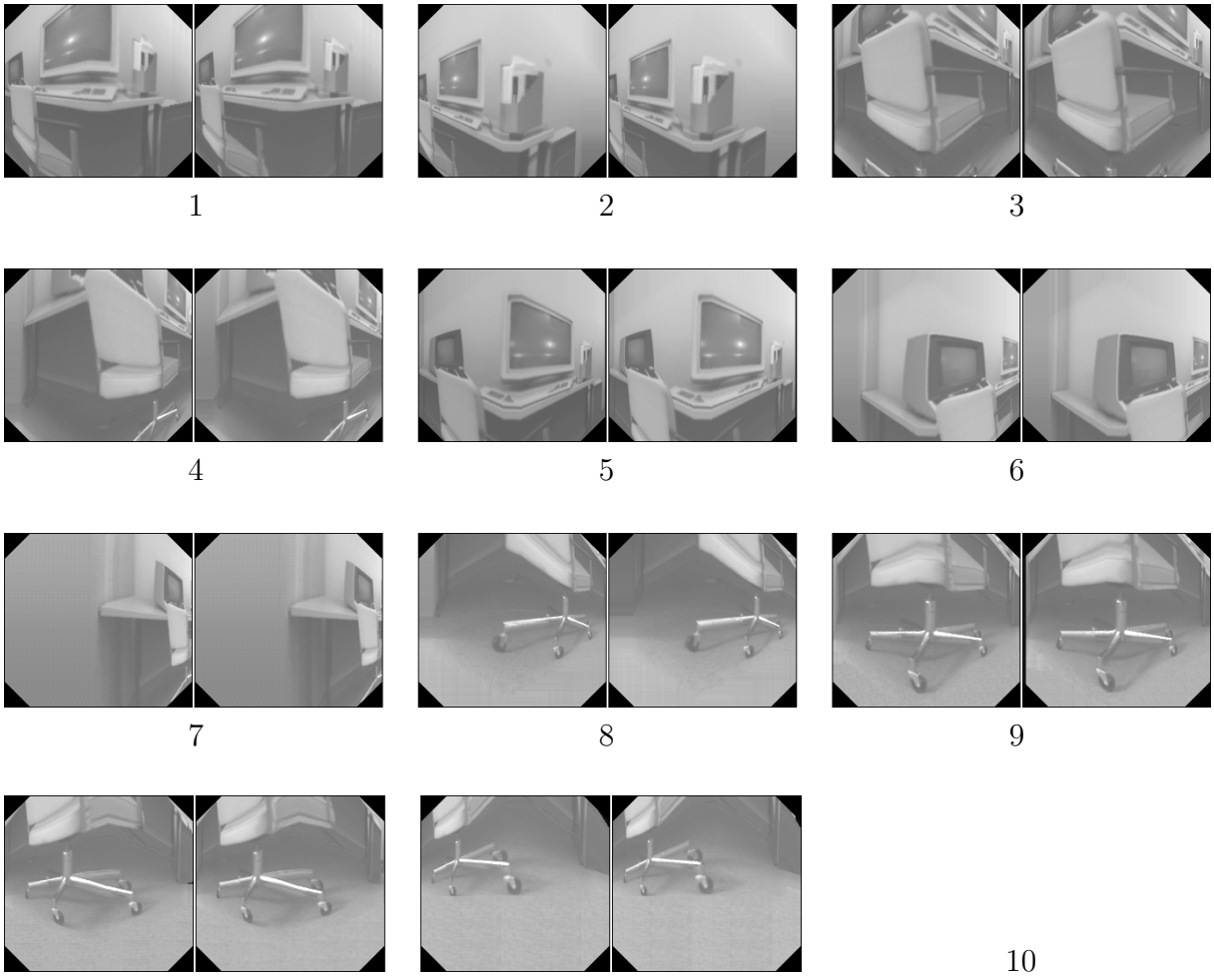

Figure 13: (a) Fixation sequence in binocular visual exploration of the office scene.

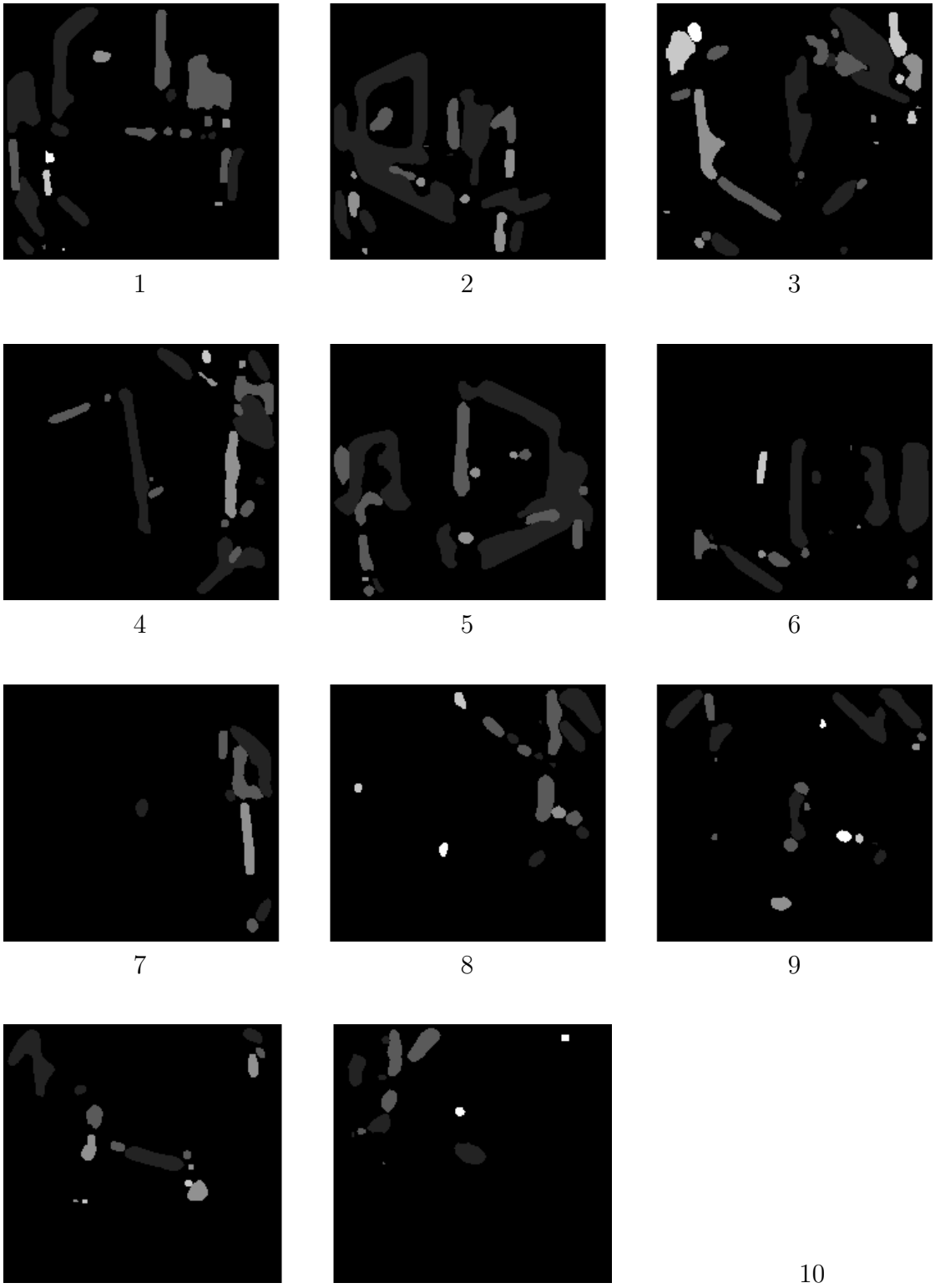

Figure 13: (b) Disparities in the RWT images.

5 Conclusion

The Reciprocal-Wedge Transform (RWT) facilitates space-variant image representation. In this paper a new RWT imaging method using the V-plane projection is presented. Based on that an RWT computational model for binocular fixation is then developed. The model provides a computational interpretation of the Panum's fusional area in relation to disparity limit in space-variant sensor space. The unique oculomotor pattern for binocular fixation observed in human system appears natural to space-variant sensing.

The RWT has its unique way of realizing foveate sensing which is shown to be beneficial for binocular fixation in active stereo vision systems. The vergence-version movement sequence is implemented as an effective fixation mechanism using the RWT imaging. The RWT simplifies the disparity computation because its variable resolution is primarily in one dimension. The horizontal displacement inflicted in stereo images due to the binocular disparity is well captured in the RWT representation.

A fixation system is presented to show the operation of various modules for camera control, vergence, version and "where-next". This paper reported our recent work in applying the RWT to camera fixation and scanpath movements in active stereo. Future research will be its applications in attention and visual exploration, e.g., in situated robots.

References

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

[2] C. Fermüller and Y. Aloimonos. Vision and action. *Image and Vision Computing*, 13(10):725–744, 1995.

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

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

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

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

[7] K. Pahlavan, T. Uhlin, and J.O. Eklundh. Dynamic fixation and active perception. *International Journal of Computer Vision*, 17(2):113–135, 1996.

[8] E. Grosso and M. Tistarelli. Active/dynamic stereo vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(9):868–879, 1995.

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

[10] N. Ahuja and A. L. Abbott. Active stereo: integrating disparity, vergence, focus, aperture, and calibration for surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1007–1029, 1993.

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

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

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

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

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

[16] F. Tong and Z.N. Li. A camera model for reciprocal-wedge transform. *Image and Vision Computing*, 14(5):339–351, 1996.

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

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

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

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

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

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