Active stereo vision and cyclotorsion

Michael R. M. Jenkin

Department of Computer Science
York University
North York, Ontario, Canada

John K. Tsotsos

Department of Computer Science
University of Toronto
Toronto, Ontario, Canada

Abstract

When a particular point is fixated by an active stereo system different portions of the world are brought into interocular alignment. This region is known as the horoptor. Through an examination of the horoptor under different viewing conditions it is demonstrated that for certain binocular tasks it is desirable to manipulate the horoptor by rotating (torquing) the cameras about their optical axes. This manipulation can be passive for operations such as stereo based obstacle detection for mobile robots, or active for active binocular heads. Techniques for both situations are presented.

1 Introduction

Stereo vision has a long history as a research topic in both the computational and biological literatures. Early stereopsis work in the computer vision field concentrated on static stereo systems utilizing parallel optical axes. More recently the computational field has begun to examine how to actively control robotic stereo sensors (stereo heads) so as to allow the vision system to 'fixate' different structures in a scene[1,3,11,13]. Most of these heads concentrate on moving the fixation point of the stereo system (the intersection point of the two optical axes) without considering the effect of different head geometries on the computational tasks that are to be performed on the acquired images. Although these effects may be safely ignored for very simple operations such as tracking a bright spot around the environment, these effects can have serious implications for more sophisticated active vision tasks such as obtaining environmental layout from active stereo. As computational stereopsis is applied under varying convergent geometry, the underlying computational tasks become more and more similar to the tasks encountered by biological stereo systems. The biological community has considerably more experience in the analysis of of convergent stereo imagery processing and it can be instructive to examine how biological systems exploit the camera/eye geometry and its relationship with image disparities.

Given a particular head geometry, three dimensional points in the world are mapped to image points in the two cameras that make up a binocular head. Labeling these cameras as 'l' and 'r', the computational task associated with localizing structure in three dimensions given the projection of the structure onto the two cameras reduces to the task of determining the correspondence between the projection of a point in one camera with its projection in the other. This is the classic *correspondence* problem and is considered to be the hard problem in stereo image processing. In the most general case the search for the corresponding feature in one camera is a two dimensional search through the entire image space of the other. Knowledge of the camera geometry can be used to limit the search to a single dimension along the epipolar line. If only a limited range of disparities needs to be searched, this search can be limited to a finite segment of the epipolar line.

In a passive stereo system, different virtual fixation points can be constructed and searches for matches performed near these points. In an active stereo system, a fixed disparity range is considered for interocular matches and different regions of three dimensional space are mapped to this disparity range by modifying the head geometry. In either case, when only a limited range of disparities are to be processed (say near zero disparity), the relationship between the range of image disparities over which correspondence takes place and the three dimensional space to which it corresponds needs to be known. If this relationship is not known then either some regions of space will not be sensed which should be or computational resources will be wasted fruitlessly searching matches in regions for which matches are not required. What effect do

1063-6919/94 $3.00 © 1994 IEEE

different head parameters have on the 3D region of real space that is localized near zero disparity?

2 The horoptor

When a scene is viewed binocularly points in the world are mapped onto different image points in the left and right eyes. In general a point will have different horizontal and vertical positions in each eye. That vertical disparities exist and have a particular structure is overlooked in much of the computational literature although there have been some notable exceptions, e.g. [8,9]. If $P = [x\ y\ z\ 1]^T$ is an arbitrary point in 3 space described in homogeneous coordinates, then a pinhole camera $i \in (l, r)$ maps P to the point $(u_i, v_i) = f_i((M_iP)_x/(M_iP)_z, (M_iP)_y/(M_iP)_z)$ where M_i is the mapping in homogeneous coordinates from a world coordinate system to a coordinate system aligned with the i'th camera and f_i is the focal length. The region of space that has the same vertical and horizontal positions in l and r is then given by $u_l = u_r$ and $v_l = v_r$. Regardless of the complexity of the geometry relating the two cameras, possible differences in focal length, and even possible camera misalignment, each of these two constraints simplifies to a term of the form

$$\sum_{j=1,2} (a_j x + b_j y + c_j z + d_j)(e_j x + f_j y + g_j z + h_j) = 0$$

The region of space having zero vertical and horizontal disparity must satisfy both of these constraints simultaneously. The structure of this *horoptor* can be very complex. Rather than examining all possible head geometries, a simpler form of the horoptor is possible by restricting head geometries to those possible with existing stereo heads.

In order that the optical axes of the two cameras intersect, most stereo heads are either constructed so that they cannot raise or lower their eyes independently or so that they control the vertical orientation of the eyes as a single logical unit. The assumption of equal tilt and focal lengths for the pinhole camera models simplifies the form of the horoptor considerably. Affix a right handed 3D co-ordinate system to the binocular head with l and r at $(-e, 0, 0)$ and $(e, 0, 0)$. As the two cameras will not be raised or lowered independently, align the y axis so that it is perpendicular to the plane formed by joining the nodal points of the two eyes with the fixation point. The positive y-axis points up, the positive x-axis points to the right and the negative z-axis points

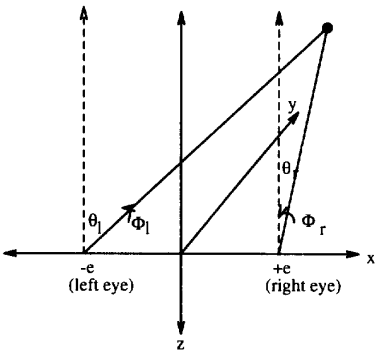

Figure 1: Coordinate system. The two eyes fixate a common point (the fixation point). The eyes deviate amounts θ_l and θ_r from looking straight ahead, and torque about their optical axes by amounts ϕ_l and ϕ_r. Note that in order for the fixation point to be in front of the head $\theta_r \geq \theta_l$.

towards the fixation point. The cameras rotate away from the $-z$-axis by amounts θ_l and θ_r and torque about their optical axis by amount ϕ_l and ϕ_r (see Figure 1). Under these assumptions the horoptor equation for zero horizontal and vertical disparity results in two constraints; both of which are of the form $ax^2 + 2hxz + bz^2 + 2gx + 2fz + c = 0$ where

$$\begin{aligned}
a = &\ \cos(\phi_l)\cos(\theta_l)\sin(\theta_r) - \cos(\phi_r)\cos(\theta_r)\sin(\theta_l) \\
h = &\ (1/2)\cos(\theta_r + \theta_l)(\cos(\phi_l) - \cos(\phi_r)) \\
b = &\ \cos(\phi_r)\cos(\theta_l)\sin(\theta_r) - \cos(\phi_l)\sin(\theta_l)\cos(\theta_r) \\
g = &\ (1/2)y(\sin(\phi_l)\sin(\theta_r) - \sin(\phi_r)\sin(\theta_l)) \\
f = &\ (1/2)e(\cos(\phi_l) + \cos(\phi_r))\cos(\theta_r - \theta_l) + \\
&\ (1/2)y(\cos(\theta_r)\sin(\phi_l) - \cos(\theta_l)\sin(\phi_r) \\
c = &\ e(e\cos(\phi_r)\cos(\theta_r)\sin(\theta_l) - \\
&\ \cos(\phi_l)\cos(\theta_l)\cos(\theta_r)) \\
&\ -y(\sin(\theta_r)\sin(\phi_l) + \sin(\phi_r)\sin(\theta_l)))
\end{aligned}$$

are the coefficients for the horizontal constraint and

$$\begin{aligned}
a = &\ \sin(\phi_l)\sin(\theta_l)\cos(\theta_r) - \sin(\phi_r)\sin(\theta_r)\cos(\theta_l) \\
h = &\ (1/2)\cos(\theta_r + \theta_l)(\sin(\phi_l) - \sin(\phi_r)) \\
b = &\ \sin(\phi_r)\sin(\theta_l)\cos(\phi_r) - \sin(\phi_r)\sin(\theta_r)\cos(\theta_l) \\
g = &\ (1/2)y(\sin(\theta_r)\cos(\phi_l) - \sin(\theta_l)\cos(\phi_r)) \\
f = &\ (1/2)y(\cos(\phi_l)\cos(\theta_r) - \cos(\phi_r)\cos(\theta_l)) \\
&\ -(1/2)e(\sin(\phi_l) + \sin(\phi_r))\cos(\theta_l - \theta_r) \\
c = &\ e(e(\sin(\phi_l)\cos(\theta_l)\cos(\theta_r) - \\
&\ \sin(\phi_r)\cos(\theta_r)\sin(\theta_l)) \\
&\ -y(\sin(\theta_r)\cos(\phi_l) + \sin(\theta_l)\cos(\phi_r)))
\end{aligned}$$

are the coefficients for the vertical constraint. These are curves of the second order and have been derived before many times using different representa-

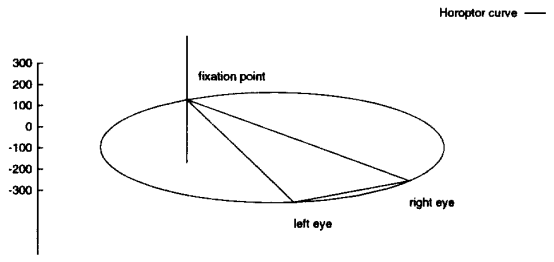

Figure 2: Horoptor curve without torsion

tions (see Helmholtz[18] for an alternative derivation). Both of these curves are potentially ellipses not aligned with the x-z axes. Regardless of the torsions both constraints are satisfied at the nodal points of the two cameras $(\pm e, 0, 0)$, and also at the fixation point $(-e\sin(\theta_r + \theta_l)/\sin(\theta_r - \theta_l), 0, -2e\cos(\theta_l)\cos(\theta_r)/\sin(\theta_r - \theta_l))$.

2.1 Zero Torsion

Now consider the (typically assumed) case of zero torsion ($\phi_l = \phi_r = 0$). The horizontal and vertical constraints simplify to the region of space described by $x^2 + (z + e\cot(\theta_r - \theta_l))^2 = e^2/sin^2(\theta_r - \theta_l)$ and $y = 0$ or $x(\sin(\theta_r) - \sin(\theta_l)) + z(\cos(\theta_r) - \cos(\theta_l)) = e(\sin(\theta_r) + \sin(\theta_l))$ This curve is plotted in Figure 2.

As has been described many times in the biological literature (see [16] or [5] for example), the horoptor curve consists of two parts, a circle lying in the plane containing the nodal points of the two eyes (known as the longitudinal horoptor), and a vertical line perpendicular to the circle (known as the vertical horoptor). The longitudinal horoptor is also known as the Vieth-Müller circle. This circle remains unchanged as the eyes fixate different points along the circle ($\theta_r - \theta_l$ remains fixed). It is also important to note that the vertical horoptor does not necessarily intersect the longitudinal horoptor at the fixation point.

When the eyes are in symmetric vergence $\theta_r = -\theta_l$ then the ideal vertical horoptor passes through the fixation point. Normally sighted human subjects show a tilt in the vertical horoptor away from the observer [18] which implies that the corresponding retinal points in humans are twisted outwards[17]. This point will be returned to later.

Consider the implications of the horoptor on classic parallel optical axes stereo algorithms. The zero-disparity region corresponds to two single lines in space. Near-zero horizontal disparities correspond to a cylindrical region which is perpendicular to the plane containing the nodal points of the eyes and the fixation point. Near zero vertical disparities limit the cylinder to regions near the $x - z$ plane and a line parallel to the y axis. The zero-disparity region most certainly does not correspond to a simple flat surface.

Another example comes from the application of static stereo cameras to the task of obstacle avoidance in mobile robotics. Typically two (or more) cameras are mounted in fixed calibrated mounts with a fixed fixation point some distance from the robot. As the robot moves different regions are brought to the fixation point, and the robot is to determine if objects exist which might require the robot to change course. Unfortunately the horoptor does not correspond with the region of space that needs to be searched for potential obstacles. Ideally the horoptor should be either vertical to detect obstacles such as walls, or coincident with the floor to detect floor anomalies. As torsion does not change the property that the fixation point has zero disparity, perhaps the local structure of the horoptor can be manipulated by choosing different eye torques for different visual tasks.

2.2 Symmetric Fixation

An examination of the horizontal disparity constraint shows that the ellipse can be made to align with the x-z axes by choosing to torque the eyes with the same magnitude. For the vertical disparity constraint the torques must be identical. However if symmetric fixation is assumed then writing $\theta_l = -\theta$, $\theta_r = \theta$, $\phi_l = -\phi$ and $\phi_r = \phi$ the horoptor simplfies to a parabola lying in the $x = 0$ plane (the vertical horoptor) and an ellipse which lies in the plane $y\sin(\theta) + z\tan(\phi) = 0$ (the longitudinal horoptor). This is shown in Figure 3. Under symmetric fixation, cyclotorsion defines the local surface slant to which zero disparity corresponds (see also [12]).

3 Passive control of cyclotorsion

Consider the task of utilizing stereo to detect floor anomalies (FAD) or obstacle avoidance for a mobile robot. Assuming no torsion, then the vertical horoptor which passes through the fixation point slopes away

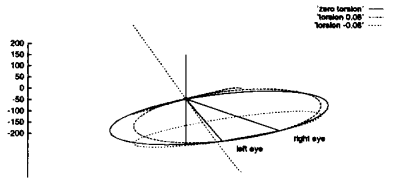

Figure 3: Horoptor curve with torsion. As the eyes torque away from zero torsion the vertical horoptor rotates away from the vertical and the horizontal horoptor moves out of the plane formed by the fixation point and the nodal points of the two eyes to meet it.

from the robot. By rotating the eyes about their optical centres this slope can be set to any value. Two values are of particular interest; setting the vertical horoptor to be perpendicular to the ground, or setting the vertical horoptor to lie on the ground plane. Setting the vertical horoptor to be vertical would be useful for obstacle avoidance tasks in man made environments. Vertical walls would be detected as large structures having spatial cohesion as the robot moves throughout its environment. Setting the vertical horoptor to lie on the ground plane would be useful for detecting anomalities on the ground plane; such as small obstacles or hollows in the ground[4]. This is the strategy that appears to be used by biological systems (see [16] for a review of this strategy in different animals and on the development of this strategy in humans).

In order to utilize stereo vision in a robotic environment for tasks like floor anomaly detection or obstacle avoidance, a reasonable amount of vision calibration must be performed. The optical axes of the camera must intersect (or intersect at infinity) in order for the interocular matches to be useful for later processing. As this calibration must be performed, and as part of this process is to calibrate for a particular torsion (typically zero), there is little additional cost associated with calibrating for a specific non-zero torsion.

4 Active control of cyclotorsion

In a binocular head such as TRISH[10] capable of actively controlling eye torsions it is possible to dynamically adjust the eye torsion so as to match the local surface tilt. It may also be possible to control torsion in software for heads without mechanically adjustable torsions although the sampling difficulties associated with sampling rotated images are well known[14]. The reduced apparent surface tilt obtained with torsional eye movements brings more of the surface within a particular range of disparities. Thus the surface can be made to appear more "fronto-parallel" and thus more easily processed with a static arrangement of disparity detectors.

There are two possible computational models for computing the 'cyclodisparity' between the left right images[6]. One mechanism would be to have a unique cyclodisparity measurement process which is specifically designed to measure gross rotational differences between the two images. A second mechanism would combine local disparity measurements into a global rotational measurement. This second form of cyclodisparity detection can be accomplished by combining local disparity estimates to solve in a least squares manner for the global image rotation.

Assume that the eyes are currently fixating some locally planer surface. Then near the centre $(0,0)$ of the image the distribution of disparities $(\delta x, \delta y)$ can be related to a local torsional rotation $\delta\phi$ by $\delta x = -y\delta\phi$ and $\delta y = x\delta\phi$. If that the eyes are fixated on some image structure, then choosing a cyclodisparity that minimizes in a least squares sense the fit near the image centre to a image rotation solves for

$$\phi = \frac{\sum_j x_j \delta y_j - \sum_i y_i \delta x_i}{\sum_j x_j^2 + \sum_i y_i^2} \qquad (1)$$

Note that this computation requires only the values of δx and δy which are the local image disparities which are typically computed by the disparity measurement process.

In order to implement the cyclodisparity measurement process, some mechanism is required to measure local disparities. A phase based interocular matching process[7] is used here but other disparity measurement processes are possible.

For rotated patterns the image rotation obtained using (1) recovers the correct cyclodisparity until the disparity grows so large that the $(\delta x, \delta y)$ values begin to fall outside the disparity range to which the disparity detectors are tuned. This is illustrated quite clearly in Figure 4.

In order to track cyclodisparities in an active environment some sort of control loop is desirable to predict the expected cyclodisparity and to smooth out small temporal variations. Many different control loops are possible. One simple approach is to

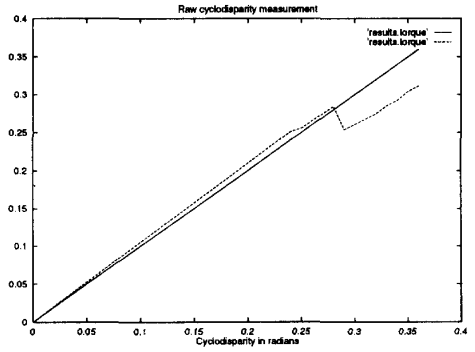

Figure 4: Recovered cyclodisparities using a phase based disparity measurement process [7] and a least squares fit. Note the slight positive bias due to the use of normal disparity rather than 2D disparity and the failure of the measurement process as the inputs begin to fall outside the detection range of the disparity measurement process.

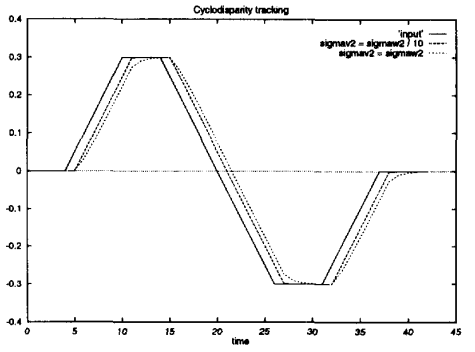

Figure 5: Active control of cyclodisparity. The horizontal scale is time while the vertical scale is the induced cyclodisparity to each eye so the total cyclodisparity is twice that shown.

model the true cyclodisparity as a constant which is corrupted with zero mean noise and to build a Kalman based control process to estimate the true cyclodisparity[2], i.e. to assume that $\phi(t)$ is simply a corrupted version of $\phi(t-1)$ and that the measurement process given in (1) returns a corrupted version $y(t)$ of $\phi(k)$, i.e. $\phi(t) = \phi(t-1) + w(t-1)$ and $y(t) = \phi(t) + v(t)$. Assuming that the noise process is well behaved, i.e. $E[w(t)] = E[v(t)] = 0$, $E[w^2(t)] = \sigma_w^2$, $E[v^2(t)] = \sigma_v^2$ and $E[w(k)w(j)] = E[v(k)v(j)] = 0$ $k \neq j$, then the Kalman estimate $\hat{\phi}(t)$ of $\phi(t)$ is given by $\hat{\phi}(t) = a(t)\hat{\phi}(t-1) + b(t)y(t)$ where $a(t) = 1 - b(t)$ and $b(t) = (\sigma_v^2 b(t-1) + \sigma_w^2)/(\sigma_v^2 b(t-1) + \sigma_w^2 + \sigma_v^2)$. This can be embedded within a simple control loop in order to recursively estimate in a least squares sense the current cyclodisparity at time t given measurements $y(0)..y(t)$.

The results of using this simple control loop to actively determine the cyclodisparity and to account for it are shown in Figure 5. The surface starts out at zero disparity and then tilts to induce a cyclodisparity of 0.6 radians in total. The surface maintains this tilt and then changes tilt until the surface induces a cyclodisparity of -0.6 radians in total. Results for two different control loops are shown. In the first $\sigma_v^2 = \sigma_w^2/10$ while in the second $\sigma_v^2 = \sigma_w^2$. In both cases the active cyclodisparity process accurately tracks the input, nullifying the induced cyclodisparity. The effect of increasing σ_v^2 is to generate a longer temporal averaging process so that the tracking is smoother but slightly delayed.

5 Discussion

Although the effect of torsional rotations on the horoptor geometry have been understood for over 100 years it is only recently that computational models of stereopsis have considered convergent stereopsis and have had to examine in depth the effects of different head geometries on the nature of the computational processes. As binocular processing is typically only performed over a small range of disparities it is important to understand the relationship between a particular head geometry, a range of disparities, and three dimensional space. This paper has derived the situation for the smallest possible range of disparities (exactly zero disparity), and it is near this region in space that most active stereo heads and processes propose to process stereo information. If stereo heads are to be used for more than just tracking single bright dots over a simple background, or to track objects at roughly eye height, the effects of the geometry cannot be ignored.

For active heads it is not possible to precompute the appropriate torque for all visual tasks. Different torques are suitable for different tasks. The active control of torsional eye movements can be used to; (i) make the best possible use of the range over which the disparity detectors operate by mapping structure in the world to the detection region of the operators, (ii) make an arbitrary surface slant appear "frontoparallel" in disparity space, and thus ideally suited to binocular processing by many stereopsis algorithms.

and (iii) replace the bottom-up search process in vision with a top-down search with explicit target knowledge which is a considerably more efficient process[15]. Active binocular systems which do not take the shape of the horoptor into account must attempt to overcome the mismatch between zero-disparity surfaces in disparity space and planar surfaces in the real world by searching large disparity regions whose size is a function not only of the shape of the 3D surface but also of the current head geometry.

References

[1] N. Ahuja and A. L. Abbott. Active stereo: Integrating disparity, vergence, focus aperature, and calibration for surface estimation. *IEEE PAMI*, 15(10):1007–1029, 1993.

[2] S. M. Bozic. *Digital and Kalman filtering.* Halsted Press, 1979.

[3] C. Brown, D. Coombs, and J. Soong. Real-time smooth pursuit tracking. In A. Blake and A. Yuille, editors, *Active Vision*, pages 155–174. MIT Press, Cambridge, MA, 1992.

[4] P. Burt, P. Anandan, K. Hanna, G. van der Wal, and R. Bassman. A front-end vision processor for vehicle navigation. In *IAS-3*, pages 653–662, Pittsburgh, PA, 1993.

[5] I. Howard. *Human Visual Orientation.* John Wiley and Sons Ltd., Chichester, NY, 1982.

[6] I. Howard. Cycloversion, cyclovergence and perceived slant. In L. Harris and M. Jenkin, editors, *Spatial Vision in humans and robots*, pages 349–366. Cambridge University Press, 1993.

[7] M. Jenkin and A. Jepson. Recovering local surface structure through local phase difference measurements. *CVGIP: IU*, 59(1):72–93, 1994.

[8] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:433–435, 1981.

[9] H. C. Longuet-Higgins. The role of the vertical dimension in stereoscopic vision. *Perception*, 11:377–386, 1982.

[10] E. Milios, M. Jenkin, and J. Tsotsos. TRISH, a binocular robot head with torsional eye movements. *Int. J. of Pat. Rec. and AI*, 7(1):51–68, 1993. Special issue on "Mobile robots, robot heads and active vision'.

[11] D. Murray, F. Du, P. McLauchlan, I. Reed, P. Sharkey, and M. Brady. Design of stereo heads. In A. Blake and A. Yuille, editors, *Active Vision*, pages 155–174. MIT Press, Cambridge, MA, 1992.

[12] K. Nakayama. Geometric and physiological aspects of depth perception. In *Proc. SPIE*, volume 120, pages 2–9, 1977.

[13] T. J. Olson. Stereopsis for verging systems. In *IEEE CVPR*, pages 55–60, New York, NY, 1993.

[14] A. V. Oppenheim and R. Schafer. *Digital signal processing.* Prentice-Hall Inc., Englewood Cliffs, NJ, 1975.

[15] J. Tsotsos. The complexity of perceptual search tasks. In *Proc. IJCAI*, pages 1571–1577, Detroit, MI, 1989.

[16] C. W. Tyler. The horoptor and binocular fusion. In D. Regan, editor, *Binocular Vision*, pages 19–37. CRC Press, Boca Raton, FL, 1991.

[17] C. W. Tyler and A. B. Scott. Binocular vision. In R. E. Records, editor, *Physiology of the human eye and visual system*, pages 643–674. Harper and Row, New York, NY, 1979.

[18] H. von Helmholtz. *Treatise on Physiological Optics.* Dover, New York, NY, 1866. First published in 1866. The 1909 edition was translated into English by J.P.C. Southall in 1924 and republished in 1962.

Acknowledgments

The second author is the CP-Unitel Fellow the Canadian Institute for Advanced Research. This research was funded by the Institute for Robotics and Intelligent Systems, a Network of Centres of Excellence of the Government of Canada, and the Natural Sciences and Engineering Research Council of Canada.