

# The horoptor and active cyclotorsion

M. R. M. Jenkin	J. K. Tsotsos	G. Dudek
Computer Science	Computer Science	Computer Science
York University	University of Toronto	McGill University
North York, Canada	Toronto, Canada	Montreal, Canada

## Abstract

*When a particular 3D point is fixated by a robotic stereo system different portions of the world are brought into interocular alignment. This region is known as the horoptor. Purposeful modifications to the binocular geometry can be used to bring different regions of three-space closer to the horoptor: Camera pan and tilt define the rough structure of the horoptor, while camera torsion can be used to change its local shape. Theoretical and empirical results suggest that for binocular vision tasks (i) it is important to understand the region of three space that contains the horoptor curve and (ii) it is possible to control this shape in an active way so as to simplify certain binocular tasks.*

## 1 Introduction

The fundamental task in binocular vision is the localization of structure in three dimensions. Given the projection of scene points into the two cameras that form a binocular head the computational task reduces to determining the correspondence between the projection of a point in one camera with its projection in the other. Knowledge of the camera geometry can be used to limit the search to a single dimension along the epipolar line. In a passive stereo system, different virtual fixation points are constructed and searches for matches are 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. 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 they correspond 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 for matches in regions for which matches are not required. What effect do different head parameters have on the 3D region of real space that is localized near zero disparity? How can head parameters be controlled so as to enhance the fit between the region of space that is to be sensed and the sensing requirements of the task at hand?

## 2 The horoptor

When a scene is viewed binocularly, points in the world are mapped onto different image points in the left and right cameras. In general a point will have different horizontal and vertical positions in each camera. The region of space that is projected onto identical image points (the horoptor) has been derived elsewhere (see [10] for a derivation using retinal coordinates and [5] for a derivation using a pinhole camera model). In either case, the region of zero disparity can be described as the solution of the intersection of two curves of the second order. When two cameras fixate a single point in space and the cameras are aligned so that their  $x$  axes lie in the same plane and their  $y$  axes are parallel, the horoptor has the classical form shown in Figure 1. As has been described many times in the biological literature (see [9] or [2] for example), the horoptor curve consists of two parts, a circle lying in the plane containing the nodal points of the two cameras (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 cameras fixate different points along the circle. It is also important to note that the vertical horoptor does not necessarily intersect the longitudinal horoptor at the fixation point.

The horoptor curve shown in Figure 1 is based on a simple numerical model that is unlikely to hold in practice. The cameras cannot be perfectly aligned vertically, nor are they pinhole cameras. In order to test

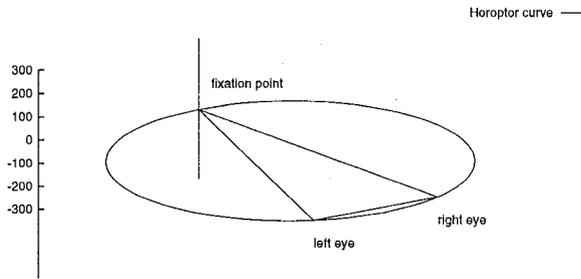


Figure 1: Horoptor curve without torsion. The two cameras fixate a single point and are not torqued about their optical axes.

the validity of the classic model, a number of different stereo rigs have been examined under various head geometries and their horoptors determined empirically. (See [6] for details on the horoptor measurement technique.) Figure 2 shows the empirically determined horoptor for a fixed stereo rig at York University with no torsion. The empirical horoptor curve shown in Figure 2 follows the theoretical circular horoptor reasonably well consisting of two parts, a (roughly) circular part which lies in the plane containing the fixation point and the two cameras, and a (roughly) vertical part near the fixation point. A second roughly vertical component near the back of the head is also admitted by the mathematical model. Note that unlike the theoretical model, the vertical horoptor is not a straight line, but rather consists of two branches which diverge slightly.

How serious an effect does torsional camera movements have on the shape of the horoptor in practice? The York head shown in Figure 2 shows some residual torsion in that the vertical horoptor tilts slightly away from the stereo head. The York head has an adjustable torsion so the induced torsion might be an artifact of a poor torsional setting in the head. Figure 3 shows the empirically determined horoptor for the MCR CIM Active Stereo Head. The MCR CIM head consists of two fixed focal length cameras mounted on two Directed Perception pan and tilt units. The head is designed to have a fixed zero torsion in each camera.

The vertical horoptor in Figure 3 tilts away from the stereo head indicating some torsional misalignment (cyclotorsion) is present. For the MCR CIM

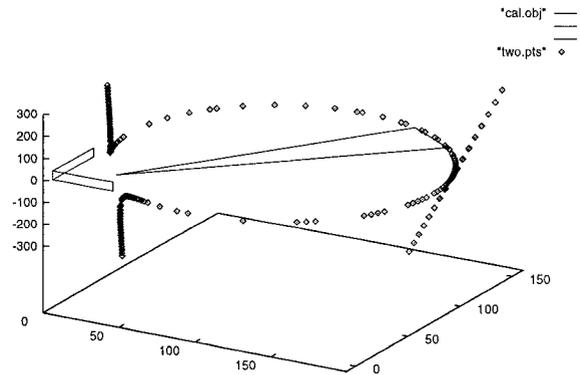


Figure 2: The empirically measured horoptor for a fixed stereo head at York University. Points determined to lie on the horoptor are marked by small diamonds. A reference object consisting of two planes is shown. This object is used to empirically measure the horoptor. The camera centres and their approximate directions of gaze are joined.

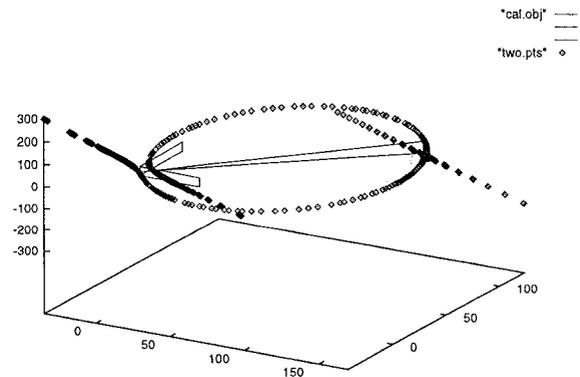


Figure 3: The empirically measured horoptor curve for the McGill University stereo head. Note that the vertical horoptor curve tilts away from the head indicating that some torque is present.

stereo head, “frontoparallel” surfaces are slanted away from the head. A large range of vertical and horizontal disparities must be searched in order for an unslanted surface to be fused by the MCRCIM head. As active stereo heads are developed which can control various aspects of the head geometry, it may prove beneficial to introduce low level control loops to correct for artifacts such as torsional or vertical misalignment in much the same way that binocular head control for vergence and version is controlled in existing systems.

### 3 Active control of cyclotorsion

In a binocular head such as TRISH[7] capable of actively controlling camera torsion it is possible to dynamically adjust the camera torsion so as to match the local surface tilt of the object being fixated. The reduced apparent surface tilt obtained with torsional camera 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.

The cyclodisparity (or rotational misalignment) between the two images can be computed in a number of different ways[3]. The approach considered here is to combine local independent disparity estimates to solve in a least squares manner for the global image rotation. Assume that the cameras are currently fixating some locally planar 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 the cameras fixate some scene point then choosing a cyclodisparity that minimizes in a least squares sense the mis-fit near the image centre to image rotation solves for

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

Note that this computation requires only the values of  $\delta x$  and  $\delta y$  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 based on [4] is used here but other disparity measurement processes could be used.

For simple 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.

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. It is beyond the scope of this paper to address all of the issues involved in designing an ideal cyclotorsion control loop here. One simple approach is to model the true cyclodisparity as a constant which is corrupted with zero mean noise and to build a Kalman filter based control process to estimate the true cyclodisparity[1]. That is 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  $\phi(t)$ , i.e. that  $\phi(t) = \phi(t-1) + w(t-1)$  and  $y(t) = \phi(t) + v(t)$ . Assuming that the noise process is well behaved, i.e. that  $E[w(t)] = E[v(t)] = 0$ ,  $E[w^2(t)] = \sigma_w^2$ ,  $E[v^2(t)] = \sigma_v^2$ , 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 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 4. 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  $\sigma_v^2$  is to generate a longer temporal averaging process so that the tracking is smoother but slightly delayed.

Camera misalignment in a binocular head can produce complex changes in the shape of the horoptor curve. By measuring and correcting for torsional misalignment it is possible to correct not for a fixed cyclotorsion, but rather for the cyclotorsion that best fits the local surface slant. This brings more of the local surface into rough alignment with the disparity detectors tuned to near zero disparity.

### 4 Discussion

Although the effect of torsional rotations on the horoptor geometry have been understood for over 100

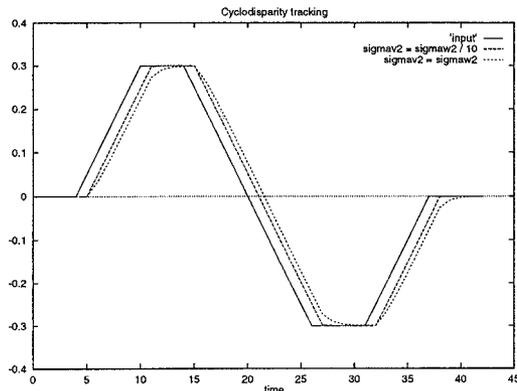


Figure 4: Active control of cyclodisparity. The horizontal scale is time while the vertical scale is the induced cyclodisparity to each camera so the total cyclodisparity is twice that shown.

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.

For active stereo heads it is not possible to pre-compute the appropriate torque for all visual tasks. Different torques are suitable for different tasks. Active control of torsional camera movements can be used to: (1) 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; (2) It can be used to make an arbitrary surface slant appear “frontoparallel” in disparity space, and thus ideally suited to binocular processing by many stereopsis algorithms; And (3) It can play a role in explicitly tuning top-down search processes using explicit target knowledge[8]. 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] S. M. Bozic. *Digital and Kalman filtering*. Halsted Press, 1979.

- [2] I. Howard. *Human Visual Orientation*. John Wiley and Sons Ltd., Chichester, NY, 1982.
- [3] I. Howard. Cyclovergence, cyclovergence and perceived slant. In L. Harris and M. Jenkin, editors, *Spatial Vision in humans and robots*, pages 349–366. Cambridge University Press, 1993.
- [4] M. Jenkin and A. Jepson. Recovering local surface structure through local phase difference measurements. *CVGIP: IU*, 59(1):72–93, 1994.
- [5] M. Jenkin and J. Tsotsos. Active stereo vision and cyclotorsion. In *IEEE CVPR'94*, pages 806–811, Seattle, 1994.
- [6] M. R. M. Jenkin and J. K. Tsotsos. Robotic stereo vision and cyclotorsion. (submitted for publication), 1994.
- [7] 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”.
- [8] J. Tsotsos. The complexity of perceptual search tasks. In *Proc. IJCAI*, pages 1571–1577, Detroit, MI, 1989.
- [9] C. W. Tyler. The horoptor and binocular fusion. In D. Regan, editor, *Binocular Vision*, pages 19–37. CRC Press, Boca Raton, FL, 1991.
- [10] 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.