CYCLOTORSION AND THE TRISH ACTIVE STEREO HEAD

Michael R. M. Jenkin[1], *Evangelos E. Milios*[1], *John K. Tsotsos*[2]

[1] Department of Computer Science, York University,
4700 Keele St., North York, Ontario, Canada,
phone: 416-736-5053, *fax:* 416-736-5872
e-mail: (jenkin,eem)@cs.yorku.ca

Department of Computer Science, University of Toronto,
6 King's College Rd., Toronto, Ontario, Canada,
phone: 416-978-3619, *fax:* 416-978-1455
e-mail: tsotsos@vis.toronto.edu

ABSTRACT

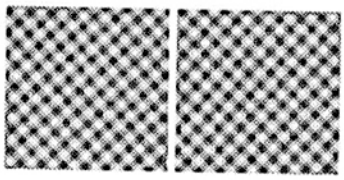
For many stereo applications fixed stereo geometry has been found to be inadequate and active stereo systems capable of manipulating the camera geometry have been developed. Head motions, and individual camera pan and tilt motions define the fixation point of an active system, while rotations of the cameras about their optical axes (torsion) defines the local slope of the zero disparity surface near this fixation point. As stereo processing typically prefers a surface slope perpendicular to the plane containing the fixation point and the nodal points of the two cameras, torsional control in an active stereo system such as TRISH can be used to improve the efficiency of the search for stereo matches and can tune the head geometry for specific object recognition tasks. This paper explores the relationship between head geometry, including non-zero camera torsions and the position and shape of the near-zero disparity surface. It develops a zero-disparity and zero-cyclo-disparity control system for tracking both the position and slant of a target using an active stereo head capable of torsional camera motions.

1. INTRODUCTION

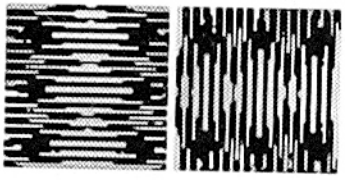
When a scene is viewed binocularly, points in the world are mapped onto different image points in the left and right cameras. In general, a point will have different horizontal and vertical positions in each camera. Assuming a pinhole camera model, then regardless of the complexity of the geometry relating the two cameras, possible differences in focal length, and even possible mis-alignment of the two cameras, the region of space that maps to zero horizontal and vertical disparity can be described as the intersection of two quadratics in x, y, and z. This region is known as the horoptor. If the two cameras fixate a point in space and if the cameras have zero torsion, then the horoptor curve has the classical form described in the literature (see [12] or [4]). The classic 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). 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, although it does so for the case of symmetric fixation.

The effect of different fixations on the near-zero disparity region can best be shown by example. Consider an active stereo head with zero torsion capable of symmetric fixation mounted on a mobile robot with the line joining the nodal points of the cameras parallel to the floor (this corresponds to the geometry of the Harvard Binocular Head [3]). When the robot fixates an obstacle at roughly the same height as the cameras the longitudinal horoptor is roughly parallel to the ground, and the vertical horoptor is roughly vertical. A large portion of the horoptor will intersect a vertical object such as a person or a wall. On the other hand, suppose that the head fixates the intersection of a vertical obstacle with the ground plane. The longitudinal horoptor lies in the plane joining the

This work was supported by NSERC Canada and the Federal Networks of Centres of Excellence IRIS Project.



(a) Input left and right images



(b) horizontal (left) and vertical (right) disparity fields

Figure 1: Computing local cyclotorsion. (a) shows the left and right inputs which are rotated with respect to each other. (b) gives the recovered horizontal and vertical disparity fields. Dark values indicate no response.

nodal points of the two cameras and the fixation point, while the vertical horoptor is perpendicular to this. The intersection of these curves with the vertical obstacle or the floor will be very small, and the obstacle will be difficult to localize binocularly.

Active manipulation of the camera torsions does not affect the zero-disparity property of the fixation point, although it does effect the local shape of the zero disparity surface[7]. Different camera torsions 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) 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 be used to replace the bottom-up search process in vision with a top-down search with explicit target knowledge which is a considerably more efficient process[11]. By actively manipulating not only the fixation point but also the local shape of the near-zero-disparity surface, the visual structure imaged by the cameras can be warped so as to be more suitable for later visual processing. In an active stereo system, control of the location of the fixation point can be performed by changing neck parameters and camera tilt, vergence and version while controlled torsional motions of the cameras can be used to manipulate the local shape of the near-zero-disparity surface.

In order to control an active stereo head it is necessary to measure visual events in the environment and to control the head based on these visual signals. For active stereo systems with a limited operational range of horizontal, vertical and orientational (cyclo-) disparity measurement, a control mechanism is required in order to keep a target within the operational range of the detectors. Low level measurement processes can be constructed to measure target disparity in both time and space[7]. Based on these low level measurements, control signals can be constructed to adapt the pose of the stereo head to the local scene structure. This is important for stereo processing as many stereopsis algorithms utilize orientation tuned filters in the earliest stages of interocular correspondence. Small orientational differences caused by non-zero surface tilts may give rise to binocular images which do not fall into the same orientation tuned mechanism operating in each of the two cameras. Although matching between different orientations interocularly is a potential solution to this problem, it is computationally expensive to consider all possible interocular orientational differences.

2. ADAPTING TO LOCAL SCENE STRUCTURE

TRISH is a seven degree of freedom stereo head. It has a single neck pan, and then each camera has three rotational degrees of freedom. Suppose that TRISH is required to fixate a particular point in space. Head pan is used so that the point is in the primary position of gaze, and the two cameras are raised and the cameras verged or panned to

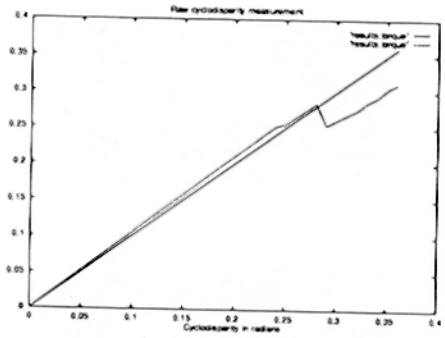


Figure 2: Recovered torques using a phase based disparity measurement process [6] and a least squares fit. Note the slight positive bias due to the use of normal disparity rather than 2D disparity and the failure of the measurement process as the inputs begin to fall outside the detection range of the disparity measurement process.

fixate the required point. Given these constraints, there are a number of mechanisms that TRISH can use to adapt to the local scene structure;

- Individual camera torsions can be controlled to adapt to the local surface tilt.
- Given potential misalignment errors in the tilt of the two cameras, an active process can be used to correct for this misalignment.
- The head can adapt to local disparity errors to more accurately verge on the structure being fixated.

There are two possible computational models for computing the 'cyclodisparity' between the left and right images[5]. 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 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 are fixated on some image structure, then choosing a cyclodisparity that minimizes in a least squares sense the fit near the image centre to image rotation solves for

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

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 based on [6] is used here but other disparity measurement processes could be used.[1] Given a pair of images the horizontal and vertical disparities can be computed using this or some other method as shown in Figure 1.

For simple rotated patterns such as the one shown in Figure 1 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 and orientation range to which the disparity detectors are tuned. This is illustrated quite clearly in Figure 2. In order to track cyclodisparities in an active environment some sort of control loop is required to predict the expected cyclodisparity and to smooth out small temporal variations. One simple approach is to model the true

[1] The technique used in [6] computes $(\delta x, \delta y)$ values which are normal disparities rather than the required 2D disparities, but for small rotational angles the difference can be ignored.

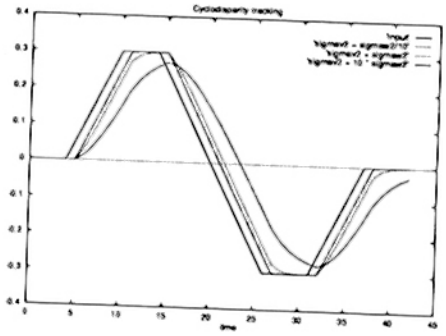


Figure 3: Active control of cyclodisparity. The horizontal scale is time while the vertical scale is the induced cyclodisparity to each eye so the total cyclodisparity is twice that shown.

cyclodisparity as a constant which is corrupted with zero mean noise and to build a Kalman 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 $\phi(k)$

$$\begin{aligned}\phi(t) &= \phi(t-1) + w(t-1)\\ y(t) &= \phi(t) + v(t)\end{aligned}$$

where $w(t)$ and $v(t)$ are the model and sensor noise respectively. Assuming that the noise process is well behaved, i.e.

$$\begin{aligned}&E[w(t)] = E[v(t)] = 0\\ &E[w^2(t)] = \sigma_w^2 \quad E[v^2(t)] = \sigma_v^2\\ &E[w(k)w(j)] = E[v(k)v(j)] = 0 \; k \neq j\end{aligned}$$

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

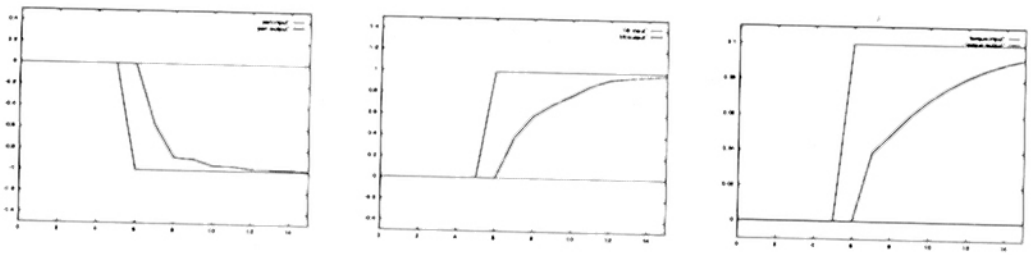
$$b(t) = \frac{\sigma_v^2 b(t-1) + \sigma_w^2}{\sigma_v^2 b(t-1) + \sigma_w^2 + \sigma_v^2}$$

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 3. 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 three different control loops are shown. In the first $\sigma_v^2 = \sigma_w^2/10$ while in the second $\sigma_v^2 = \sigma_w^2$, and in the third $\sigma_v^2 = 10\sigma_w^2$. In the first two cases the active cyclodisparity process accurately tracks the input, nullifying the induced cyclodisparity, while the time constant of the third case is sufficiently long that the tracking does not fully complete in the plot shown here. The effect of increasing σ_v^2 is to generate a longer temporal averaging process so that the tracking is smoother but delayed.

Now consider the problem of adapting to vertical misalignment between the left and right images. As with the case for torsion, we can combine local disparity measurements into a global vertical misalignment measure and then drive the tilt motors to correct for this. Assume that the cameras are currently fixating some locally planar surface, the distribution of disparities $(\delta x, \delta y)$ can be related to a local vertical misalignment $\delta\alpha$ by $\delta x = 0$, and $\delta y = \delta\alpha$. A control loop identical to the one developed for torsion above using a Kalman filter to model a corrupted constant signal can be constructed to integrate measurements over time.

A similar mechanism can be built to adapt to local variations in disparity. Again assuming ideal shifts, the local horizontal misalignment $\delta\theta$ is related to the distribution of disparities by $\delta x = \delta\theta$ and $\delta y = 0$.



(a) Horizontal Disparity (b) Vertical Disparity (c) Torsional Disparity

Figure 4: Combined adaptation to horizontal, vertical and torsional disparities. The same control mechanism was used but with different time constants.

How then can these three different control mechanisms be combined in order to adapt to all three error signals? One mechanism is to run all three control mechanisms in parallel but to introduce different time constants so as to reduce the effect of interactions between the different controls. Figure 4 shows all three control mechanisms in operation simultaneously, as the input is misaligned vertically, horizontally, and torsionally. Here all three control loops have the same value of σ_w^2, but σ_v^2 (torsion) = $10\sigma_v^2$ (tilt) = $10\sigma_v^2$ (vergence). The Horizontal and Vertical scales are in subsampled pixels (a unit of one corresponds to an eight pixel disparity in the full image), while torsional disparity is measured in radians per camera, so 0.1 corresponds to about 12° of rotational difference.

3. DISCUSSION

Torsional eye movements have been ignored by most stereo researchers in computer vision. 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. If stereo heads are to be used for more than just tracking dots over a simple background the effects of the geometry cannot be ignored.

Passive geometry tasks such as obstacle avoidance or floor anomaly detection can be greatly simplified by choosing to rotate the eyes about their optical centres. As this rotation can be relatively small for appropriate baseline, eye height and fixation distance choices, it is a simple modification to existing binocularly guided mobile robots. But this simple change can result in considerable computational savings. A small modification to the geometry of the horoptor replaces the search through a large disparity region for anomalies to a much smaller region near the horoptor. It is interesting to note that experimental determination of the horoptor shows that humans have a predisposition for a non-zero cyclodisparity, and that different binocular species also show this predisposition[2]. Binocular biological systems seem to torque their eyes so as to drive the vertical horoptor to the ground plane for distant fixation distances.

For active heads it is not possible to precompute the appropriate torque for all visual tasks. Different torques are suitable for different tasks, and active stereo heads can be designed to actively control for torsion, in a manner similar to the process of controlling vergence or tilt.

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.

4. REFERENCES

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

[2] M. L. Cooper and J. D. Pettigrew. A neurophysiological determination of the vertical horoptor in the cat and owl. *J. Comp. Neurol.*, 184:1–26, 1979.

[3] N. Ferrier. The Harvard binocular head. In *SPIE Conference on AI X: Machine vision and robotics*, pages 2–13, Orlando, FL, 1992.

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

[5] 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.

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

[7] M. Jenkin and J. Tsotsos. Active stereo vision and cyclotorsion. In *IEEE CVPR'94*, pages 806–811, Seattle, 1994.

[8] M. Jenkin, J. Tsotsos, and G. Dudek. The horoptor and active cyclotorsion. In *IAPR Conference on Pattern Recognition*, Jerusalem, 1994.

[9] 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'.

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

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

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

[13] 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.