On the use of trajectory information to assist stereopsis in a dynamic environment

Michael R. M. Jenkin
Department of Computer Science, York University
Toronto, Ontario, Canada

Abstract

If stereopsis is to be used in a dynamic environment, it makes little sense to re-compute the entire representation of disparity space from scratch at each time step. One simple approach would be to use the results from the current solution to "prime" the algorithm for the next solution. If three dimensional trajectory information was available, this information could be used to first update the previous solution, and then this updated solution could be used to "prime" the algorithm for the following stereo pair. Recent work[4, 5] has demonstrated that it is possible to measure such trajectory information very quickly without complex token or feature extraction. This paper demonstrates how raw disparity measurement made by this earlier technique can be integrated into a single trajectory measurement at each image point. A mechanism is then proposed that updates a stereopsis algorithm operating in a dynamic environment using this trajectory information.

1 Introduction

For the most part, the application of computers to the task of stereopsis has been restricted to the static analysis of images, although some dynamic stereopsis algorithms have been developed. A static pair of images is presented to the cameras, and then some process for determining scene heights is applied to the two images. Even when applied to what are inherently dynamic tasks such as pick and place, or mobile robotics tasks, a static algorithm is applied over and over again. In light of this, some efforts have been made to embed stereopsis within temporal processing. For example, optical flow can be used to limit possible correspondences between the left and right images[12, 13], and point based correspondences can be considered in space-time, rather just in space or time[3, 7]. Another possibility is to use active vision techniques to reformulate stereopsis as an active task. Olson and Potter[9] have used Cepstral filtering and deconvolution to drive camera vergence movements. A similar philosophy drives this paper, but at a more local level. Rather than obtaining a single (or small number) of values that indicate the relative vergence and version that will be required to gaze onto structure in a scene, the technique presented in this paper obtains local measurements of three dimensional motion (stereomotion), and uses these measurements coupled with an earlier solution to the stereo problem to *prime* the stereo algorithm for the following time step. More explicitly, given a solution to the stereo problem at time t and a measurement of the local trajectory from time t to $t + \delta t$, a local approximation to the solution at time $t + \delta t$ can be computed. This first approximation can then be refined by the static stereopsis algorithm when it is applied at $t + \delta t$.

A complete implementation of a time-varying stereopsis algorithm is beyond the scope of this paper.

This paper addresses how trajectory information can be recovered and suggests a mechanism by which disparity estimates can be updated. A full algorithm which recovers surface and object descriptions in a dynamic environment is the subject of ongoing research.

2 Measuring a Unique Trajectory

Previous results have established that;

- The disparity between two bandpass signals can be determined by computing the relative phase difference between the two signals[6, 11]. This measurement can be performed directly from Sine and Cosine Gabor filtered versions of the input images.
- Space-time Sine and Cosine Gabor filters are available that are also selective for particular monocular image velocities[1].
- By combining the above two results it is possible to build detectors that measure particular left-right combinations of image velocity[4, 5]. These detectors are a very simple model of the stereomotion detectors found in biological binocular systems[2, 10].

The three dimensional trajectory detectors respond to a particular left-right velocity pair. In order to obtain a unique trajectory at a particular image point the responses of these detectors must be combined. Suppose that a number of detectors with the same speed and spatial frequency specificity are operating at a given point in an image but with different preferred trajectories. Then if structure passes through that point, each detector will respond with a projection of the true trajectory onto the particular trajectory the detector is tuned to. The response of the detectors can be modeled as $R(x) = A_0 + A_1 \cos(x + A_2)$ where $R(x)$ is the response of the x'th detector (tuned for three dimensional trajectory x). Least squares can be used to fit the detector responses to $R(x)$ and obtain estimates for A_1 (the strength of the response), and A_2 (the trajectory). Note that the responses cannot be fit to a least squares solution directly, and that $R(x)$ must be reformulated as $R(x) = \alpha + \beta\cos(x) + \gamma\sin(x)$ where $A_0 = \alpha$, $A_1^2 = \beta^2 + \gamma^2$ and $A_2 = \tan^{-1}(-\gamma/\beta)$. By recovering A_1 and A_2 we have a measure of the three dimensional trajectory and the strength of the response by pooling a number of detectors. One interesting special case of this formulation occurs when only four detectors are used tuned to motion in the frontoparallel plane to the left (R_1), motion directly away from the observer (R_2), motion in the frontoparallel plane to the right (R_3), and motion directly towards the observer (R_4). α, β and γ can then be recovered as $\alpha = (R_1 + R_2 + R_3 + R_4)/4$, $\beta = (R_1 - R_3)/2$, and $\gamma = (R_2 - R_4)/2$. β and γ then correspond to the opponent mechanism suggested in earlier papers[4, 5]. They also correspond to using two orthogonal energy detectors[1] and then using the projection of the two detectors to obtain the three dimensional trajectory. Note, however, that fitting the response to a sinusoid to obtain the trajectory is a more general technique and can easily be extended to obtain a trajectory measurement with more than four (or even four different) detectors.

Once a unique trajectory has been obtained by a given velocity channel, the outputs of the different velocity channels must be combined in order to obtain a final three dimensional trajectory at each point. As each velocity channel is returning both a direction and a strength for the preferred velocity, a winner take all approach based on the strength of the response is used to choose at each point the appropriate trajectory and velocity. A threshold is used to ignore detectors which have very low strength responses.

In order to show the promise of the technique, the trajectory measurement technique presented here is embedded within a very simple stereopsis algorithm. This implementation is designed only for exposition of the use of trajectory information and is not designed to be a complete stereo algorithm. Suppose that D_i are a collection of disparity demons, and that these demons tile the 'x' dimension. The demons are all tuned to the same spatial frequency, and each demon is tuned to zero disparity. At each time step each demon computes a new lock and certainty value, and updates its disparity lock position to the detected disparity. Associated with each demon is a trajectory measurement, and for the demons which have a valid lock on structure, their home disparity is updated by the disparity component of the recovered trajectory. Note that this is a very simple technique. Much more sophisticated methods are available for the disparity demons when used in a static environment[8].

3 A simple simulation

Figure 1 shows the left and right view of a random dot target oscillating in depth about zero disparity. Detectors were applied to the image selective for three dimensional trajectories with speeds of 1, 2 and 4 pixels/frame, and with directions of "towards", "away", "to the left", and "to the right". The response of the four detectors are then fitted to a Sine curve and a strength R_1 and phase (trajectory) R_2 is computed at each point. Four such families were constructed with speeds of 1, 2, and 4 pixels/frame. For each demon at each time step, the detector with the highest response was chosen, and a single three dimensional trajectory was obtained. For points with no acceptable trajectory measurement, a speed of zero was assumed, and the disparity and positional component of the recovered trajectory is shown in Figure 2. Integration of the three trajectory channels resulted in a very strong and easily recognizable trajectory measurement in depth. It is important to remember that each trajectory measurement has been made independently, and no mechanism has been applied to smooth the responses. Such a mechanism could be used to identify the incoherent responses obtained in the position component. Finally the trajectory information is used to update the home position of the disparity detectors. The recovered disparity of the detectors is shown in Figure 3. Once again, raw responses are plotted and no process has been used to smooth the responses or to identify responses of low certainty.

4 Discussion

This paper proposes that in order to avoid the computational expense involved in recomputing the stereo solution from scratch at each time step, it would be more useful to use the previous solution plus known trajectory information as a first approximation to the new solution. Recent results have shown that with very little computational expense (low level image filtering which can be performed in parallel, and often by specialized hardware), raw trajectory measurements can be made. These trajectory demons are modeled after computational process suggested by Beverley and Regan[2], and they have many similar properties. By fitting the responses of the detectors to a Sine wave, it is possible to pool detectors located at the same spatial position with similar speed and frequency tuning into a single response. A number of such pooled responses can be aggregated together in order to make a final measurement of the trajectory that structure with a particular spatial frequency has when it passes near a particular point in three space. This final trajectory measurement can be used in a Taylor series expansion of the known disparity at a given time, to predict the disparity at a slightly later time. This prediction can then be used to prime the more expensive disparity task.

References

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Optical Society of America A*, 2(2):284–299, 1985.

[2] K.I. Beverley and D. Regan. Evidence for the existance of neural mechanisms selectively sensitive to direction of movement. *J. Physiology*, 193:17–29, 1973.

[3] M. R. M. Jenkin. Tracking three-dimensional moving light displays. In *1983 ACM SIGART-SIGGRPAH Workshop on Motion*, pages 66–70, Toronto, 1983.

[4] M. R. M. Jenkin and A. Jepson. Measuring trajectory. In *IEEE Workshop on Visual Motion*, pages 31–37, Irvine, California, 1989.

[5] M. R. M. Jenkin and A. Jepson. Response profiles of trajectory detectors. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1617–1622, 1989.

[6] M. R. M. Jenkin, A. D. Jepson, and J. K. Tsotsos. Techniques of disparity measurement. Technical Report RBCV-TR-87-16, Researches in Biological and Computational Vision, Department of Computer Science, University of Toronto, 1987.

[7] M. R. M. Jenkin and J. K. Tsotsos. Applying temporal constraints to the dynamic stereo problem. *Computer Vision, Graphics, and Image Processing*, 33:16–32, 1986.

[8] A. Jepson and M. R. M. Jenkin. The fast computation of disparity from phase differences. In *CVPR 89*, pages 398–403, San Diego, California, 1989.

[9] T. J. Olson and R. D. Potter. Real-time vergence control. In *CVPR 89*, pages 404–409, San Diego, California, 1989.

[10] T. Poggio and W. H. Talbot. Mechanisms of static and dynamic stereopsis in foveal cortex of rhesus monkey. *J. Physiology*, 315:469–492, 1981.

[11] T. D. Sanger. Stereo disparity computation using gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[12] A. M. Waxman and J. H. Duncan. Binocular image flows: steps towards stereo-motion fusion. Technical Report CAR-TR-119, Centre for Automation Research, University of Maryland, May 1985.

[13] A. M. Waxman and S. S. Sinha. Dynamic stereo: Passive ranging to moving objects from relative image flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):406–412, 1986.

These are views of a 1D (spatial) time varying stereogram. Intensity is encoded as height. A central region oscillated in disparity.

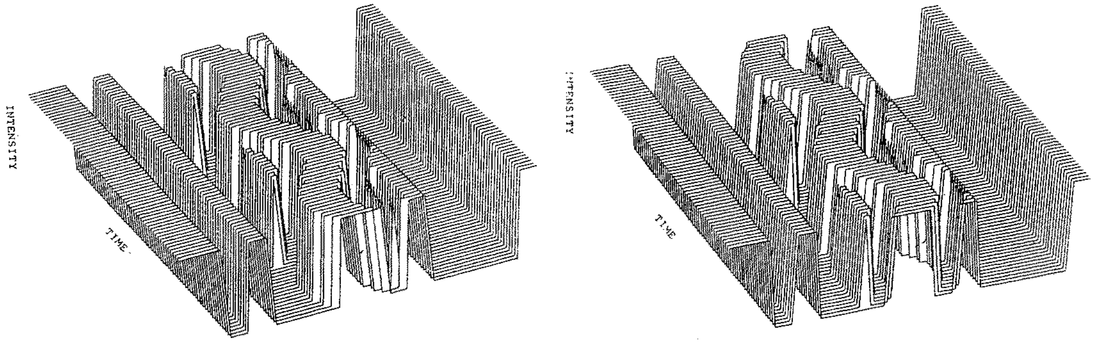
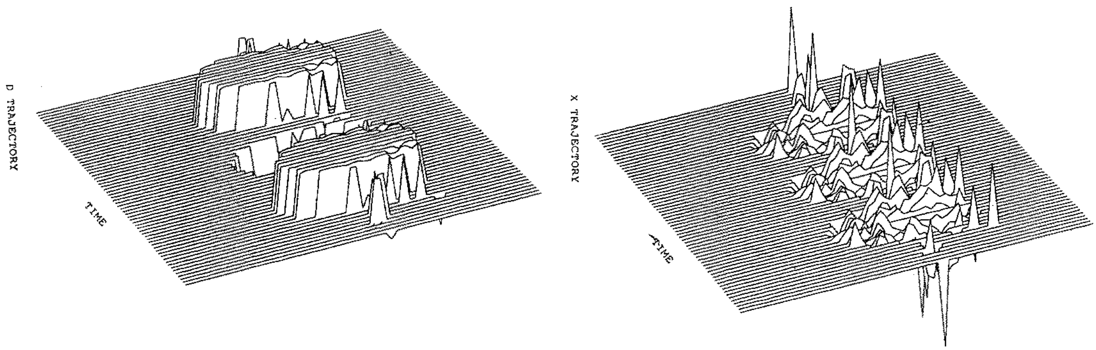


Figure 1: Left and right views of random dot stereogram



(a) Disparity component (b) Position component

Figure 2: Recovered three dimensional trajectories. Speed is encoded as height.

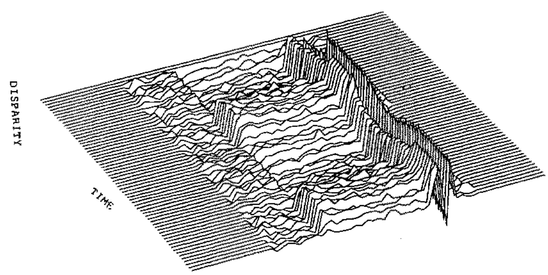


Figure 3: Disparity Demons response: Disparity is encoded as height