Tracking Three Dimensional Moving Light Displays

Michael Jenkin

Department of Computer Science
University of Toronto

ABSTRACT

A method is presented for tracking the three-dimensional motion of points from their changing two-dimensional perspective images as viewed by a nonconvergent binocular vision system. The algorithm relies on a general smoothness assumption to guide the tracking process, and application of the tracking algorithm to a three-dimensional moving light display based on Cutting's[1] Walker program as well as other domains are discussed.

Evidence is presented relating the tracking algorithm to certain beliefs about neurophysiological structures in the visual cortex.

1. Introduction

One of the most basic problems in computer vision is that of determining the correspondence between two or more sets of points. Such a correspondence may be required to follow an object through time, or to locate the same object in different views of a scene. Various algorithms have been suggested as solutions for each of these problems. Rashid [9] tracked points in a moving light display (MLD) by predicting each point's expected location. Marr and Poggio [3] developed an algorithm for determining depth from a grey scale stereo view of a scene. In our case it was required to track a three-dimensional MLD through time by fusing the two views of a scene: In effect we had to solve both problems.

A MLD is defined by Rashid [9] to be, "a sequence of binary images representing points of one or more moving objects in an actual or synthesised scene". A three dimensional moving light display is a MLD in which the lights are located in three dimensions rather than two. A MLD is a useful starting point for analysis. There is a large reduction in the number of data points from grey scale images, and object detection and recognition in a MLD does not generally reduce to frame by frame analysis. Johansson found that human subjects asked to identify the contents of a single frame are frequently unsuccessful [5].

A program was written based on Cutting's Walker program [1] to generate a three dimensional MLD similar to the two dimensional MLD described by Rashid [9]. Instead of generating a two-dimensional perspective view of the synthetic walker, the program was designed to generate a stereo view of the figure by producing two perspective views, as viewed by two eyes a fixed distance apart. A display was produced that could be viewed using a stereo viewer (fig. 1). In the display the lights were joined up to make the object more visible. An algorithm was required which would be able to match up the left and right views of a scene, and which at the same time would track the three-dimensional points through time.

Algorithms using constraints such as Ullman's rigidity constraint [12], Hoffman's planarity assumption [4], and Webb's rotation and translation assumption [13] all restrict the type of object to be tracked to be rigid, or to be composed of rigid parts. We required an algorithm that would not be restricted in either the type of object that could be tracked, nor in the types of motion the object could exhibit. Instead of basing the algorithm on assumptions of the above type, we assumed that the object's motion would obey a general smoothness assumption.

The smoothness assumption was based upon certain beliefs from Gestalt Psychology. The law of proximity, and rule of good continuation[6], suggested that we organize objects into higher groups if they are close together, and that we prefer objects to be smooth rather than to end abruptly. The principal of least action[11] suggested that when an object is perceived as moving we tend to perceive it as moving along a path that in some sense is the shortest, simplest or most direct. A temporal extension of these laws formed the basis of our smoothness assumption.

The tracking algorithm assumed a basic smoothness assumption, a given point would be relatively unchanged from one frame to the next. In particular that

1) The location of a given point would be relatively unchanged from one frame to the next.
2) The scalar velocity, or speed of a given point would be relatively unchanged from one frame to the next.
3) The direction of motion of a given point would be relatively unchanged from one frame to the next.

As these constraints treat each moving point independently they do not restrict either the type of motion exhibited by the object nor do they restrict the type of object to be tracked.

In order to simplify the problem it was assumed that at all times, all of the points were visible by both eyes. This precluded the possibility of occlusion of data points. It was also assumed that the three dimensional position and velocity of the points were known at some initial time.

2. Tracking Algorithm

Let (x,y,z) be a point in a right handed co-ordinate system. If we view this point using a non-convergent binocular vision system whose "eyes" are located at (-e,0,f) and (+e,0,f) looking towards infinity in the -z direction; then the left and right eyes will view a point $P=(x,y,z)$ as images in the x,y plane as

$$x_r = \frac{(x-e)*f}{f-z} \tag{1a}$$

$$y_r = \frac{y*f}{f-z} \tag{1b}$$

$$x_l = \frac{(x+e)*f}{f-z} \tag{2a}$$

$$y_l = \frac{y*f}{f-z} \tag{2b}$$

Where f is the focal length, the system baseline length is 2e, and (x_l,y_l) and (x_r,y_r) are the left and right eye views respectively[2].

Provided that a point (x,y,z) is visible by both the left and right eyes, then the three-dimensional co-ordinate of the point can be reconstructed as

$$x = \frac{e*(x_r+x_l)}{x_l-x_r} \tag{3a}$$

$$y = \frac{2*e*y_r}{x_l-x_r} \tag{3b}$$

$$z = f - \frac{2*e*f}{x_l-x_r} \tag{3c}$$

by solving for x, y and z in equations (1) and (2).

Let L_j, R_k be left and right eye co-ordinates of point j in the left eye, and k in the right eye. We will denote the three dimensional point which can be constructed from L_j and R_k using (3) as [j,k].

Let φ_l and φ_r be functions mapping points in the left and right images respectively from time t to time $t+1$. We seek functions $\varphi_l : n \to n$ and $\varphi_r : n \to n$, such that φ_l and φ_r are one-to-one and onto, and that the point P_i at time t has location $[\varphi_l(i), \varphi_r(i)]$ at time $t+1$. We can now state the problem more formally.

Given $P_1(t)$ to $P_n(t)$, the three dimensional location of the points at time t, $V_1(t)$ to $V_n(t)$, the vector velocity of the same n points at time t, and L_1 to L_m and R_1 to R_m (where L_i does not necessarily correspond to R_i), the left and right eye views of the points at time $t+1$, we seek one-to-one and onto mappings φ_l and φ_r such that P_i has location $[\varphi_l(i), \varphi_r(i)]$ at time $t+1$.

Let $w(i,j,k)$ be a disparity function measuring the difference between $P_i(t)$ and the three-dimensional point constructed from L_j and R_k at time $t+1$. Define $w(i,j,k)$ as

$$w(i,j,k) = d(i,j,k) + v(i,j,k) \tag{4}$$

where d and v are

$$d(i,j,k) = \frac{distance(P_i(t),[j,k])}{\sum_{l,r} distance(P_i(t),[l,r])}$$

$$v(i,j,k) = \frac{\left| V_i(t) - \frac{([j,k]-P_i(t))}{time} \right|}{\sum_{l,r} \left| V_i(t) - \frac{([l,r]-P_i(t))}{time} \right|}$$

where the sums over l and r take on all valid combinations for the left and right eye views, *time* is the time between snapshots, *distance* is the Euclidean distance between two points and $|a|$ denotes the length of the vector a.

A point $[l,r]$ is not a valid point if the absolute value of the y coordinate of L_l minus the y coordinate of R_r is larger than a given tolerance, where the tolerance is related to the resolution and noise of the input data. This is clear since from equations (1b) and (2b) it must be true that $y_l = y_r$ for the point $[l,r]$ to be valid.

The d component weights $w(i,j,k)$ by the displacement between P_i and the candidate point $[j,k]$. This corresponds to the first of the basic smoothness assumptions. The v component weights $w(i,j,k)$ by the difference in the vector velocity V_i and the vector velocity over the next time interval if the point $[j,k]$ is chosen. This corresponds to the second and third basic smoothness assumptions.

It remains to determine the correspondence functions φ_l and φ_r given $w(i,j,k)$. We wish these functions to have the property that

$$\Omega = \sum_{i=1}^{n} w(i,\varphi_l(i),\varphi_r(i)) \tag{5}$$

is minimized.

For each point P_i at time t there are at most n^2 possible matchings $[l,r]$ at time $t+1$. This gives rise to $(n!)^2$ total possible matchings. Searching through all possible mappings is not practical. We approximated the solution by applying the greedy heuristic. We chose the n smallest values of $w(i,j,k)$ to construct the mappings φ_l and φ_r, only excluding an entry if it would violate the one-to-one and onto properties of the φ_l and φ_r entries already chosen.

3. Experimental Results

The tracking algorithm was applied to three different sets of input data. The first set of data was that of a synthetic walker (fig. 1). Two other sets of input data were generated. The first produced images of colliding hexagons (fig. 3), while the second produced an image of a translating, rotating disk (fig. 5). In all of these figures certain points were joined in order to make the object more visible. Each figure represents a multiple exposure of some of the frames of the film produced.

A graphic display program was written to show the results of the tracking algorithm. The values of the correspondence function were displayed, as well as the value of Ω for each frame. The results from tracking the synthetic walker (fig 2.), the colliding hexagons (fig. 4), and the rotating, translating disk (fig. 6) were very informative. They demonstrated that the tracking algorithm was capable of tracking objects that obeyed the general smoothness assumption, and that the value of Ω was a useful measure of the uncertainty of the matching found by the algorithm.

The input to the tracking algorithm was generated so that the correct mappings φ_l and φ_r would be the identity mappings. We note that this was the case for the synthetic walker (figs. 1 and 2). It was found that from arbitrary orientation the tracking algorithm was capable of tracking the figure correctly.

The results obtained with the colliding hexagons were as expected. At the point of collision the algorithm fails - the collision violates the smoothness assumption. It was interesting that the value of Ω reaches a maximum for this frame. It was also found that the value of Ω reached a maximum while tracking the synthetic walker when the body parts passed nearest each other.

The final set of input data was that of a rotating translating disk (fig 5.). It was found that for simple rotation below a specific angular velocity the tracking was perfect. At higher angular velocities some points were tracked as rotating backwards, while others were tracked correctly (fig. 6). This is an effect that is found in human perception, quickly rotating objects appear to rotate backwards at certain speeds. Fast acceleration violates the smoothness assumption, and the tracking algorithm fails.

4. Discussion

Orban [7] performed experiments with humans on the perception of moving dots during saccades. On the basis of his experiments he put forward the following hypothesis

1) Two different detection mechanisms exist within the visual system, one dealing with velocity (and direction), the other with amplitude (and direction) of displacement.
2) The input to both of these movement channels is the movement of retinal images.
3) In the displacement channel an evaluation of entering information is made as a function of an expected value.

In later work with cats, Orban [8] discovered that all neurons exhibiting velocity specificity, or amplitude sensitivity or specificity, were also direction specific. As a result of this research Orban hypothesised that we perceive velocity or amplitude of motion as a vector quantity. Our tracking algorithm has these properties. In particular we used the result that we perceive velocity as a vector quantity, rather than as a scalar speed and direction, to construct the v term of the weight function.

In Orban's model for the first steps of cortical elaboration of movement perception [8], he proposed that the simple cells of area 18 of the cat's visual cortex receive as input the velocities of moving points on the retina from the Y-neurons of the lateral geniculate nucleus. He identified with these cells the ability to analyze retinal images, at low velocities, in terms of direction, velocity and length of movement. This is exactly the function of our tracking algorithm.

Certain functional properties of these simple cells are also exhibited by our tracking algorithm. Both fail to track at higher velocities, the simple cells do not respond [6], and our algorithm fails. Orban found that the simple cells did not respond to random dot patterns, or moving square patterns [8]. This effect was also exhibited by the tracking algorithm. When presented with a sequence of random dots in three space, the tracking algorithm tracks but produces a large value of Ω.

The tracking algorithm makes no attempt to model the low level structure of the simple cells. Rather we have attempted to model the three dimensional functional equivalent of the simple cells, and we have found that properties exhibited by these cells are also exhibited by our algorithm.

5. Conclusions

The tracking algorithm was used to track the three dimensional moving light displays generated by the programs mentioned above (figs. 1, 3, 5). For those frames in which the general smoothness assumption was satisfied tracking was performed correctly. In developing the algorithm it was found that it was necessary to use both location and vector velocity weighting to find the correct matching. It was found that in the three-dimensional case predictive functions such as Rashid's [9] were unable to track smoothly moving points correctly.

Experimentation with different weightings for the distance and velocity components revealed that the two components were equally important in determining a correspondence. Normalization of the two factors was used to weight the two components equally.

As a side effect the tracking algorithm computes the value of Ω, the total weight per frame of the chosen correspondence. The value of Ω is a measure of how inaccurate a given correspondence is. For example, in the case of the colliding hexagons (fig 2.), there is a sharp increase in the value of Ω during the frame of collision (fig 3.). This value would be very useful to higher level processes. The value of Ω could be compared against a threshold value, and could be used to "turn off" the algorithm whenever the smoothness constraints are violated.

The use of a greedy heuristic to construct the matching has both advantages and disadvantages. As a heuristic it is possible that although the weight function predicts the correct matching, the greedy heuristic will miss it. It will however predict the best matches first. Regan et al. [10] suggested that velocity ratios would be a useful technique for detecting objects moving at velocities greater than that of the rest of the scene. This processing could be performed quickly, allowing fast response to potentially dangerous objects. The greedy heuristic produces similar results but for different reasons. Matches will

be found for objects that differ significantly from the rest of the scene. For example, in the process of tracking the walking figure the head and feet were usually tracked first; these features stood out from the rest of the figure. Similarly points which have different velocities from the rest of the image will be matched early in the processing.

6. References

[1] Cutting, James E., A Program to Generate Synthetic Walkers as Dynamic Point-light Displays, Behavior Research Methods and Instrumentation, 10(1), 91-94.

[2] Duda, Richard O. and Hart, Peter E., Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

[3] Marr, D. and Poggio, T., Cooperative Computation of Stereo Disparity, Science 194, 1976, 283-287.

[4] Hoffman, D. D. and Flichbaugh B. E., The Interpretation of Biological Motion, Massachusetts Institute of Technology A.I. Memo No. 608, 1980.

[5] Johansson, G., Visual Perception and Biological Motion and a Model for its Analysis, Perception and Psychophysics 14, 2, 201-211, 1973.

[6] Koffa, K., Principles of Gestalt Psychology, Harcourt, Brace and Co., New York, 1935.

[7] Orban G. A., et al., Movement Perception During Voluntary Saccadic Eye Movements, Vision Res. 13, 1343-1353, 1973.

[8] Orban G. A., Visual Cortical Mechanisms of Movement Perception, Katholieke Universiteit Leuven, Belgium, 1975.

[9] Rashid, R. F., Towards a System for the Interpretation of Moving Light Displays, IEEE PAMI PAMI-2, November 1980, 574-581.

[10] Regan D. et al., The Visual Perception of Motion in Depth, Scientific American, Vol 241, No. 1, July 1979, 136-151.

[11] Shepard, Roger N., and Cooper, Lynn A., Mental Images and their Transformations, MIT Press, Cambridge, 1982.

[12] Ullman S., The Interpretation of Visual Motion, MIT Press, Cambridge, 1979.

[13] Webb, J. A., Structure from Motion of Rigid and Jointed Objects, Proceedings of the 7th IJCAI-81, Vancouver.

7. Acknowledgements

I would like to thank John Tsotsos and Bernd Neumann for their helpful suggestions and criticisms during the preparation of this paper.

Financial support was gratefully received from a NSERC Summer Research Award and from a contract administered by the Defense Research Establishment Atlantic.

8. Figures

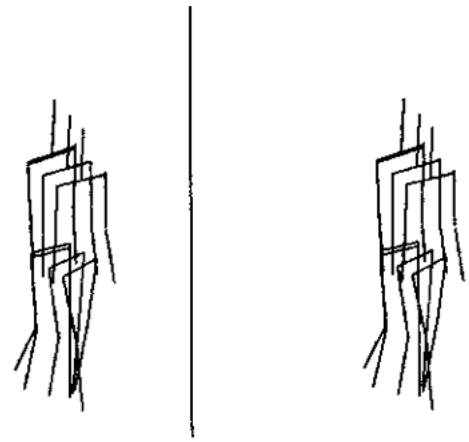

Figure 1

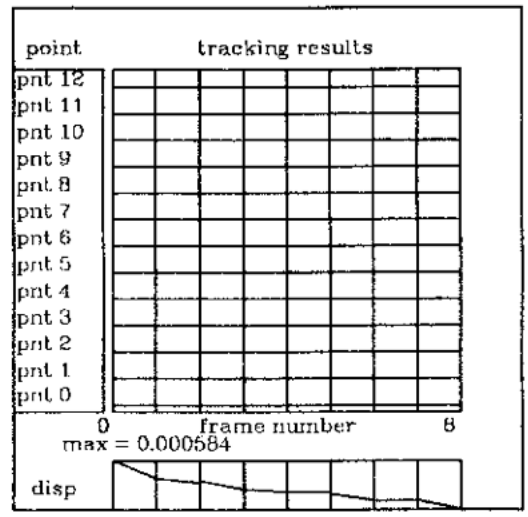

Figure 2

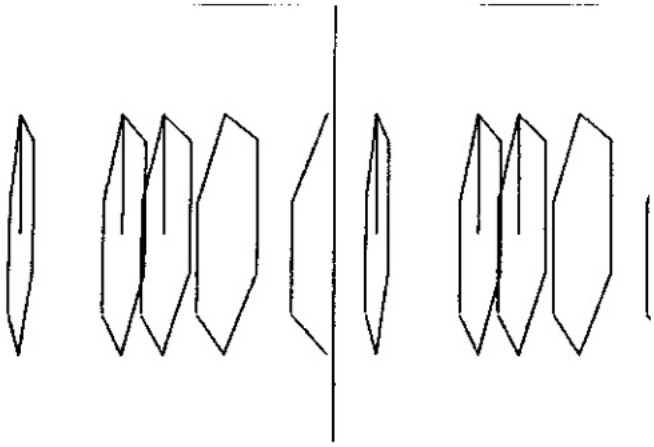

Figure 3

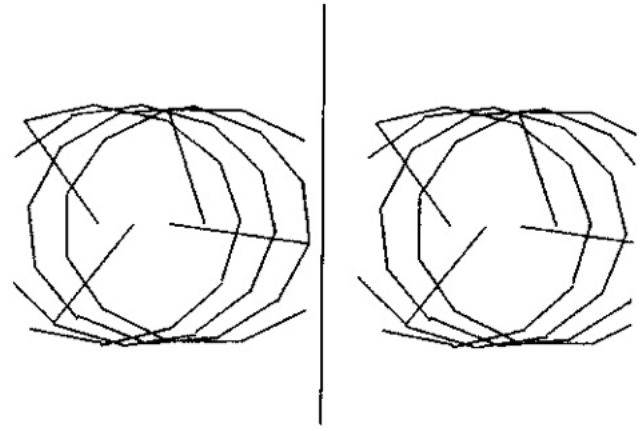

Figure 5

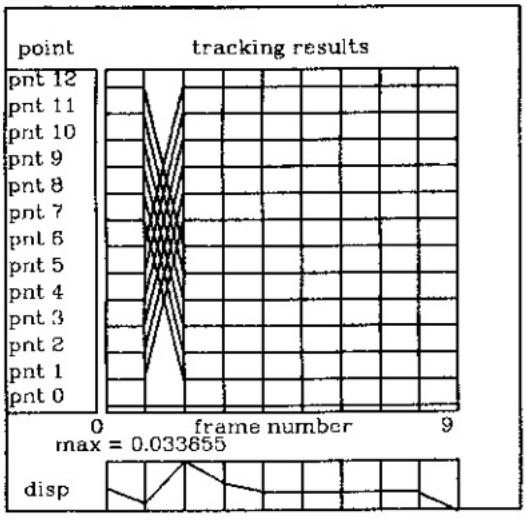

Figure 4

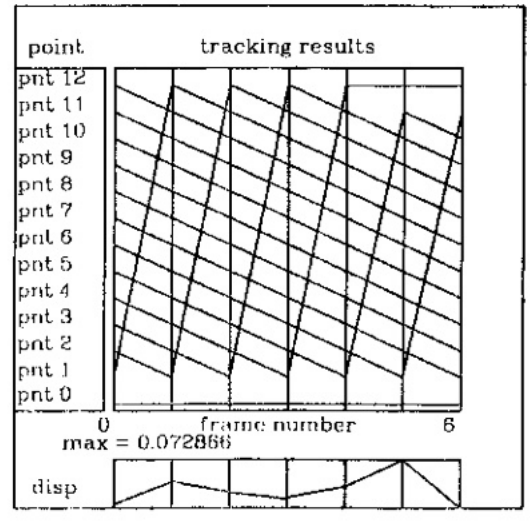

Figure 6