Dynamic Fixation of a Moving Surface Using Log Polar Sampling

Hilary Tunley and David Young
School of Cognitive and Computing Sciences
University of Sussex
Brighton, England BN1 9QH.

Abstract

We describe the development and testing of a first-order motion estimation algorithm which maintains accurate fixation of features on surfaces undergoing three-dimensional motion, and determines the local affine motion parallax. The accuracy of the first-order flow estimation is much improved by the use of log-polar sampling. We investigate the contribution of fixation to this accuracy using synthetic flow, and demonstrate the performance on affine tracking in real image sequences.

1 Introduction

A great deal of attention has recently been given to active vision – both controlling sensor motion and taking account of that motion during processing. For active vision, operating as part of a closed-loop control system, the fast, reliable estimation of a few important parameters may be more relevant than the fine-grained analysis of more traditional computer vision. Useful parameters are likely to relate to the sensor's motion relative to surfaces in the environment, and the orientation of those surfaces, but will not necessarily constitute a full three-dimensional map of the environment or a full specification of sensor translation and rotation. For moving sensors in dynamic environments, fixation (the tracking of an image feature to stabilise its retinal location) is clearly an important mechanism, contributing to robustness by reducing the demands on low-level processing mechanisms.

Active vision opens up the possibility of exploiting non-uniform image sampling, and foveal vision in particular. Despite the fact that many animals have evolved foveal visual systems, in which acuity varies across the retina, foveal vision has played little part in the development of computational vision. There appear to be three main reasons for this: firstly, computer vision has been dominated by technologies based on image sampling using uniform rectangular arrays of pixels; secondly, foveal vision, by introducing a region of high acuity, increases the need for fixation of salient features; thirdly, there is an apparent increase in algorithmic complexity for some low-level vision operations, such as motion estimation.

BMVC 1994 doi:10.5244/C.8.57

The first obstacle may be overcome by the introduction of special-purpose hardware, such as that developed for log-polar sampling by IMEC [7], and in the meantime is adequately addressed by resampling conventional images in software, an operation of low computational cost. As pointed out above, fixation is of value in its own right in the active vision paradigm, and so the need for fixation need not hinder the use of foveal vision. Indeed, various researchers have shown that the use of an active vision system can reduce the complexity of visual processing tasks by supplying constraints which are not present when passively viewing the environment [1] [2] [4]. Finally, it is increasingly clear that the complexity of non-uniform sampling is more apparent than real, and that the information carried by images can be made more accessible when the acuity is appropriately matched to the structure of the optic array.

We demonstrate here one way in which fixation and motion parameter estimation can be integrated for a particular type of foveal vision system, using a simple but reliable computational scheme.

The first-order spatial derivatives of optic flow – dilation, shear and rotation – supply useful information concerning the motions and orientations of surfaces relative to an observer [10], and are likely to be useful for active vision. (Dilation and shear are independent of sensor rotations, whilst image rotation is affected only by sensor rotation about the line of sight and not by pan and tilt movements.) Log-polar sampled images (LSI's) are a natural representation for estimating first-order motion, since the pixel spacing is proportional to distance from the origin, as is the flow speed for pure first-order flow. In addition, the high concentration of samples close to the origin of the LSI provides a natural way to focus processing on regions corresponding to coherent surfaces. We have shown that these factors allow more accurate first-order flow estimation using log-polar sampling than using the conventionally-sampled image (CSI) directly [8]. Here, we use simultaneous estimation of first-order and zero-order flow in the LSI to carry out fixation, both in the sense of tracking a feature using simulated eye movements, and in the sense of following the affine deformations of an image region during motion relative to an approximately planar surface.

In Section 2 of this paper we define the log-polar mapping and derive a simple method for first- and zero-order motion parameter extraction. Section 3 discusses implementation issues, and in Section 4 we examine the accuracy with which first-order motion parameters can be extracted using the LSI, both with and without active fixation, both applied to synthetic and real optic flow. Finally, Section 5 draws some conclusions.

2 Theory

2.1 Log Polar Sampling and First-Order Motion Analysis

Log-polar mapping involves a circularly-symmetric sampling strategy which has a much higher density at the centre (or fixation) point, and which decreases linearly with radial distance (Figure 1). Instead of representing points on the image in terms of (x, y) coordinates, a point is indexed by a logarithmic distance from the centre, ξ, and an angle, γ (after [7]), where:

$$\xi = \log_a \rho - p \quad \text{and} \quad \gamma = q\eta \tag{1}$$

Here, p and a are constants which determine the sampling used by the log-polar mapping, and (ρ, η) are the polar coordinates of the point (i.e. $x = \rho \cos \eta$ and $y = \rho \sin \eta$). For more information on LSIs see [8].

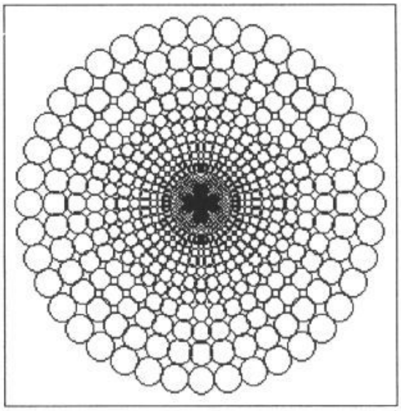

Figure 1: Layout of pixels in log-polar sampling

First-order flow includes only terms linear in image coordinates. Though higher-order terms are generally present in real flow fields, the zero- and first-order flow is a good approximation in regions for which the depth variation is small compared to the depth, and where the scene structure can be approximated by a plane. Rather than attempting to determine a flow vector at each pixel, which is computationally expensive and ill-conditioned, we estimate the six parameters of the zero- and first-order flow field directly from spatial and temporal grey-level gradients, applying the brightness constancy equation – an approach related to [9].

We represent the flow by the Taylor expansion at the origin:

$$\mathbf{v}(\mathbf{r}) = \mathbf{v}_0 + \mathbf{T}\mathbf{r} + \text{ higher order terms} \tag{2}$$

where $\mathbf{r}$ is image position, $[x\ y]^T$, $\mathbf{v}_0$ is the zero-order flow at the origin, and $\mathbf{T}$ is the deformation rate tensor of first derivatives of the flow at $\mathbf{r}$, which characterises first-order flow:

$$\mathbf{T}_{ij} = \frac{\partial \mathbf{v}_i}{\partial \mathbf{r}_j} \quad \text{and} \quad \mathbf{T} = \left[\begin{array}{cc} D + S_1 & S_2 - R \\ S_2 + R & D - S_1 \end{array} \right] \tag{3}$$

D is dilation, R, rotation, and S_1 and S_2 are the components of shear (i.e. $S_1 = S \cos 2\theta$ and $S_2 = S \sin 2\theta$ where θ is the orientation of the axis of expansion measured relative to the x axis).

If fixation can be maintained and the fixated surface is relatively smooth, then $\mathbf{v}_0 = 0$ and near the origin the first-order flow terms dominate to give:

$$\mathbf{v}(\mathbf{r}) \simeq \mathbf{Tr} \tag{4}$$

In this case, $|\mathbf{v}|$ is proportional to the distance from the origin, $|\mathbf{r}| = \rho$, which suggests that a good strategy for image sampling is to make the spatial sample separation also proportional to $|\mathbf{r}|$, as in the LSI. The image motion between successive frames will then be a constant fraction of the sample separation at every point in the image. This provides the best basis for integrating information across a region of the image.

The optic flow field in log-polar coordinates can be related to the conventional form (5) and combined with (2) to create an expression for flow, up to first-order, in log polar coordinates (6):

$$\begin{bmatrix} \dot{\xi} \\ \dot{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \gamma}{\partial x} & \frac{\partial \gamma}{\partial y} \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{1}{\rho} \begin{bmatrix} \frac{\cos\eta}{\log a} & \frac{\sin\eta}{\log a} \\ -q\sin\eta & q\cos\eta \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} \dot{\xi}\log a \\ \frac{\dot{\gamma}}{q} \end{bmatrix} = \begin{bmatrix} \frac{1}{\rho}\cos\eta & \frac{1}{\rho}\sin\eta & 1 & 0 & \cos 2\eta & \sin 2\eta \\ -\frac{1}{\rho}\sin\eta & \frac{1}{\rho}\cos\eta & 0 & 1 & -\sin 2\eta & \cos 2\eta \end{bmatrix} \begin{bmatrix} v_{x_0} \\ v_{y_0} \\ D \\ R \\ S_1 \\ S_2 \end{bmatrix} \tag{6}$$

Combining (6) with the brightness constancy assumption, $\frac{dI}{dt} = 0$, rearranging and choosing $q = \frac{1}{\log a}$ gives:

$$\dot{I} = -\begin{bmatrix} \frac{\partial I}{\partial \xi} & \frac{\partial I}{\partial \gamma} \end{bmatrix} \begin{bmatrix} \dot{\xi} \\ \dot{\gamma} \end{bmatrix} = -\begin{bmatrix} \frac{g_x}{\rho} & \frac{g_y}{\rho} & g_\xi & g_\gamma & g_u & g_v \end{bmatrix} \begin{bmatrix} v_{x_0} \\ v_{y_0} \\ D \\ R \\ S_1 \\ S_2 \end{bmatrix}$$

$$\begin{aligned} g_\xi &= \frac{1}{\log a}\frac{\partial I}{\partial \xi}; \quad g_\gamma = q\frac{\partial I}{\partial \gamma}; \quad g_x = g_\xi\cos\eta - g_\gamma\sin\eta; \quad g_y = g_\xi\sin\eta + g_\gamma\cos\eta; \\ g_u &= g_\xi\cos 2\eta - g_\gamma\sin 2\eta; \quad g_v = g_\xi\sin 2\eta + g_\gamma\cos 2\eta \end{aligned} \tag{7}$$

One equation is obtained for each sample, and it is straighforward to solve this overdetermined system by least-squares and estimate the vector of unknowns $[v_{x_0}\ v_{y_0}\ D\ R\ S_1\ S_2]^T$.

2.2 The Role of Fixation

Although the method described above provides estimates of the zero-order flow (v_{x_0} and v_{y_0}) as well as of the four first-order flow parameters, a large zero-order flow will cause the system to fail, because of the fine-grained sample spacing near the origin. A CSI has the complementary problem: it is well suited to measuring uniform flow but not first-order flow, because in this case the speed varies greatly across the image. The solution for the LSI is to use fixation to minimise the zero-order flow. This requires us to feed back the v_{x_0} and v_{y_0} estimates to a tracking system, leaving the first-order flow dominant and allowing accurate estimation of its parameters. Note that no equivalent solution is available to the problem of flow estimation using a CSI. In the sections that follow, we demonstrate the use of zero-order flow feedback to improve first-order flow estimation using the LSI.

3 Implementation

3.1 Log Polar Sampling

In the absence of dedicated hardware, conventional images were resampled using software. The strategy used, like that of other software-based LSI research (e.g. [5]), needs to balance processing speed with fidelity. Central LSI regions, with small sample spacings, require bilinear interpolation between the four nearest neighbours in the original image, whilst towards the periphery, where the samples are much further apart, simple averages of the grey-levels in a roughly circular region of the image, centred on the log-polar sampling point are used. Strategies are switched at an intermediate point. For the experimental work reported here, the innermost LSI ring had a radius of 1 pixel in the conventional image, whilst the outmost radius varied between 30 and 60 pixels. There were 100 radial samples (rings), whilst the number of angular samples (wedges) was chosen to satisfy the condition for circular sampling regions, and was typically about 150.

3.2 Controlling Motion: Parameter Extraction Accuracy

Simulated optic flow was used to test parameter extraction. Each image sequence was generated by subjecting a single conventional image to successively larger affine transformations. This created realistic motion statistics within each frame, but with known flow field parameters. Each base image was deformed by incrementing the parameters ($v_{x_0}, v_{y_0}, D, R, S_1$ and S_2) over a total of 5 or 10 frames, and resampled to form LSI sequences, with an outer radius of 60 pixels. The relative root-mean-square (RMS) errors between the actual (affine) parameters and those extracted were then calculated to provide a measure of accuracy. This error measure compares the average variation between extracted and input parameter values to the input value to provide a percentage error measure. Tests were carried out using broad-band artificial textures – binary random dot images blurred with Gaussians of 2 and 3 pixels – and a range of deformation rates from 0.01 to 0.1 per frame for D, R and S, (θ held constant) whilst v_{x_0} and v_{y_0} varied from 1 to 10 pixels. All the parameters were varied simultaneously, except for an experiment designed to assess the effect of translation alone, to ensure complex motions were examined with a realistic degree of interference between motion parameters.

In the non-fixation case the sampling was centred constantly at the same point in each warped image, regardless of the image shift (v_{x_0} and v_{y_0}), whilst the fixation case fed back these zero-order components to effectively null the shifting effect using an iterative technique. For each new frame an initial fixation point was calculated from the shift parameters determined for the previous frame. The image was then resampled, centred on that initial point and new shift parameters calculated which were, in turn, used to move the fixation point again (where necessary). This process was repeated until a fixation point was reached in which any further movement increased the calculated shift parameters.

4 Results

4.1 Static Sensor Versus Active Fixation

The results are presented in graphical form in Figures 2 to 6 for each parameter in turn. The degree of deformation is represented on the x axis, whilst the y axis indicates the normalised RMS error. Within the graph legends texture2(3) refers to random texture with Gaussian blur using a standard deviation of 2(3), and f5(10) refers to the number of frames over which the RMS is calculated. For the non-fixation case the error in the extracted parameters increases sharply with increased image motion whilst the fixating case has a significantly lower error rate over a far wider distortion range, as expected. However, it is necessary that the v_{x_0} and v_{y_0} components are accurate enough to give an initial estimate of the direction in which to perform the tracking and this level of accuracy starts to reduce above image shifts of about 3.0 pixels/frame.

We investigated the performance of the method on pure translational flow, in order to assess the range of image speeds over which fixation could be maintained in the absence of first-order deformation (Figure 7). The reduction in error due to an absence of first-order motion is apparent up to shifts of between 4 and 7 pixels/frame, dependent upon the image spatial frequencies. However, at higher speeds, tracking breaks down abruptly. This is probably due to the minimisation technique used to determine the feedback for active fixation which, like most such methods, becomes increasingly less accurate as the minimisation surface becomes flatter – a situation which occurs as the image shifts become larger and the parameters more random. Interestingly, the effect of spatial frequency on motion perception ability observed in this experiment mimics that obtained in recent psychophysical tests using similar blurred random textures and simple motion shifts [6], which revealed that increased spatial blurring increases the threshold of displacement shift in apparent motion perceivable by the human visual system. The variation in tracking accuracy between images, which is also affected by the positioning of the LSI centre – in that centering on a uniform image region supplies a large proportion of image sample points with uninformative input signals – is reduced in the presence of more complex motions where parameter extraction starts to breaks down at around 3-4 pixels/frame. Therefore, as long as the temporal sampling rates are kept high enough to ensure image shifts are kept at a couple of CSI pixels/frame at the point of fixation this tracking scheme is acceptably accurate.

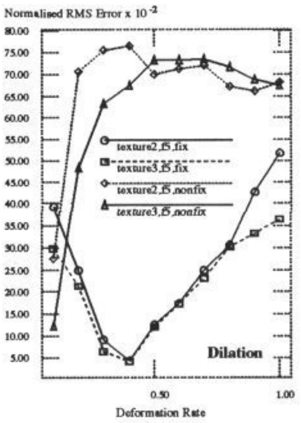

Figure 2: Dilation Extraction Accuracy

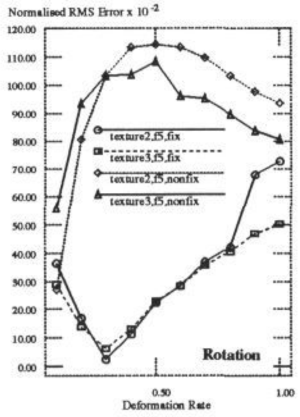

Figure 3: Rotation Extraction Accuracy

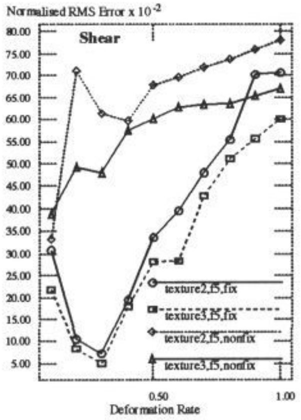

Figure 4: Shear Extraction Accuracy

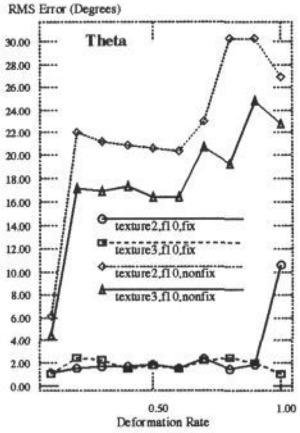

Figure 5: Theta Extraction Accuracy

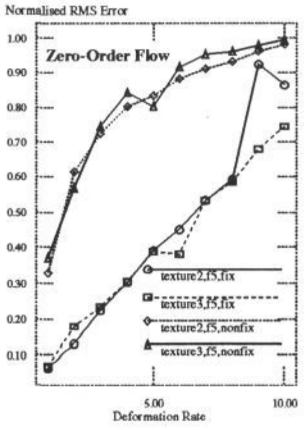

Figure 6: Shift Extraction Accuracy

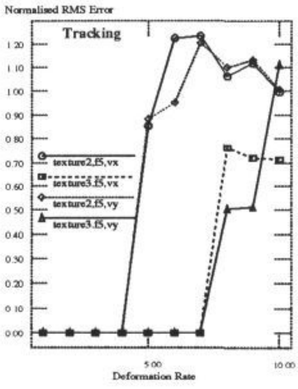

Figure 7: Tracking Accuracy

4.2 Real Image Sequences

The tests discussed above are applied to pure affine motion. For a more realistic test of performance the method was also applied to real image sequences consisting of multiple surfaces at different depths, to determine how well such an approximation holds in reality. Figures 8, 9 and 10 show the performance of the method applied to image sequences obtained from a moving camera.

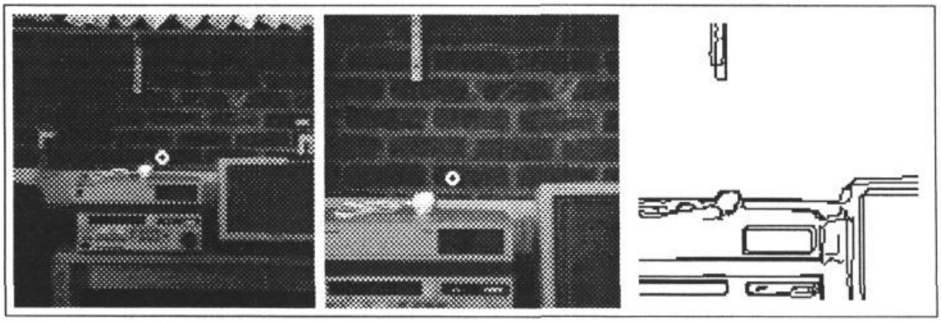

Figure 8: Example of Fixation during Approach

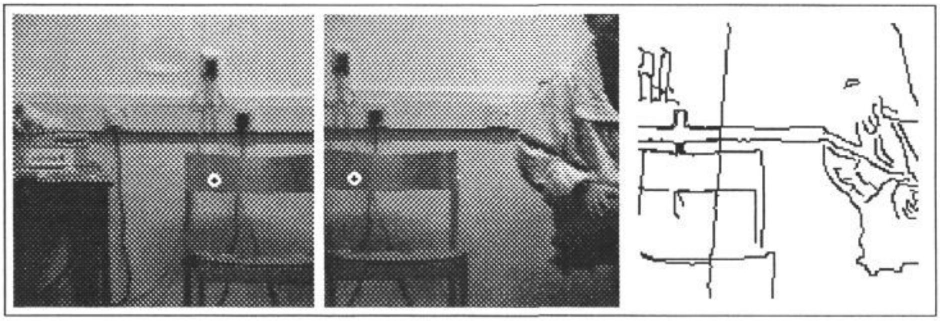

Figure 9: Example of Fixation during Panning

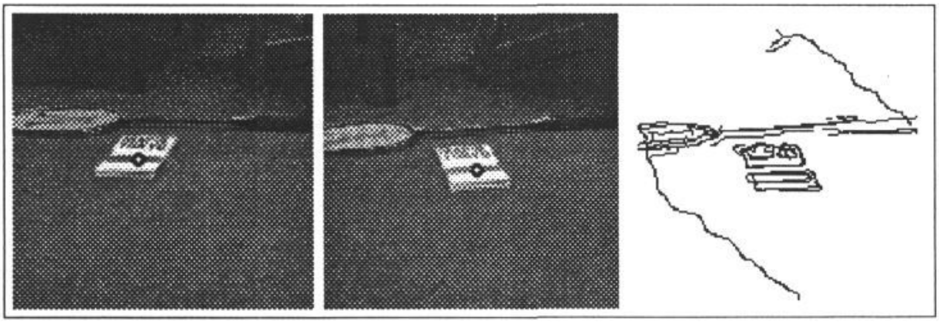

Figure 10: Example of Fixation during Sideways Translation

In each case, the images shown are the first and last images of a 20-image sequence obtained with a smoothly moving, but not accurately controlled, camera. The images were processed using zero-order flow feedback to control fixation, as described above. The dots (added after processing) indicate the initial and final fixation points. Only successive pairs of frames were processed together, so that errors in tracking were cumulative over the sequence. In each case, the tracking accuracy is good. The three sequences are dominated by dilation, translation and shear image motions respectively, corresponding to camera approach, panning and sideways translation.

The right-hand frames show the image edges from the last frame of the sequence, superimposed on the edges of the first frame, transformed according to the estimated first-order flow parameters. For the first two sequences, the estimated parameters from frame to frame were combined to give a single affine transformation which was applied to the first image. In the dilating case, it can be seen that whilst good, the fit is not perfect, but this is inevitable since the visible objects do not lie in a single plane, and so an affine transformation can only be an approximation. In the panning case, overlap is only possible for a small part of the image; this is the part to the left of the slanting line, which represents the transformed right-hand side of the first frame. In this region, the match is very close. The transformation from the first to the last frame of the shearing sequence is not well approximated by an affine transformation, and so the first frame was successively deformed by each set of flow parameters to obtain the edges shown superimposed at the right. The slanting lines are the distorted edges of the first frame.

In all three examples, almost the same set of edges is visible in the first and final frames, so the edge maps show a good degree of matching. We emphasise that the flow was estimated only for successive image pairs, so that the match of the transformed first frame to the last frame includes accumulated errors from 19 estimated transformations.

An accurate quantitative test on a real image sequence requires calibrated camera motion; we hope to carry this out in the future.

5 Conclusions and Future Work

The sample spacing in a log-polar sampled image is matched to the speed of first-order optic flow: both are proportional to distance from the origin. Provided that fixation is implemented to null the zero-order flow, the LSI spacing is therefore matched to the dominant optic flow component for local image regions. Since increasing sample spacing with speed clearly makes sense, the good performance of the LSI is expected, and was demonstrated by direct comparison with a conventional sampling approach in [8]. Here, we have explicitly shown the contribution that fixation makes to the performance, and we have demonstrated the ability of the method to carry out affine tracking on real sequences of images.

The future aims of this work include the integration of first-order motion over time and space in an effort to obtain qualitative information of both sensor motion and the motion of independent objects, and a more detailed analysis of how the stimuli used effects the performance of the sensor. Also, in most real-world situations, motion segmentation is required to ensure that image regions belonging to coherently-moving surfaces are analysed correctly. A motion segmentation scheme based on first-order modelling is thus also under development.

Acknowledgements

This work was supported by grants from the ESRC and the Joint Council Initiative in Cognitive Science/HCI.

References

[1] Aloimonos, J., Weiss, I. and Bandopadhay, A., Active vision, *International Journal of Computer Vision* **2**, 333–356, 1988.

[2] Bandopadhay, A. and Ballard, D.H., Egomotion perception using visual tracking. *Computational Intelligence* **7**, 39–47, 1991.

[3] Cipolla, R. and Blake, A., Surface orientation and time to contact from image divergence and deformation. *Proc. of the European Conf. on Computer Vision (ECCV)*, 187–202, 1992.

[4] Fermüller, C. and Aloimonos, J., The role of fixation in visual motion analysis. *International Journal of Computer Vision* **11(2)**, 165–186, 1993.

[5] Jain, R.C., Bartlett, S.L. and O'Brien, N., Motion stereo using ego-motion complex logarithmic mapping. *IEEE Trans. PAMI* **9(3)**, 356–369, 1987.

[6] Mather, G. and Tunley, H., Temporal filtering enhances motion detection. *Proc. of the European Conf. on Visual Perception*, and *Perception* **22**, p31, 1993.

[7] Tistarelli, M. and Sandini, G., On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Trans Pattern Analysis and Machine Intelligence* **15(4)**, 401–410, 1992.

[8] Tunley, H. and Young, D.S., First order optic flow from log-polar sampled images. *Proc. of the European Conf. on Computer Vision (ECCV)*, Stockholm, 1994.

[9] Werkhoven, P. and Koenderink, J.J., Extraction of Motion Parallax in the Visual System I. *Biological Cybernetics*, **63**, 185–191, 1990.

[10] Young, D.S., First order optic flow. *Cognitive Science Research Paper* **313**, University of Sussex, 1993.