Egomotion Estimation with a Biologically-Inspired Hemispherical Camera

Will Maddern and Gordon Wyeth
School of Engineering Systems, Faculty of Built Environment and Engineering
Queensland University of Technology
{w.maddern, gordon.wyeth}@qut.edu.au

Abstract

Biological inspiration has produced some successful solutions for estimation of self motion from visual information. This paper describes the advantages of egomotion estimation on a spherical image plane with respect to translational and rotational motion decoupling. The paper presents a novel 6 DoF egomotion estimation method for spherical cameras based on the image interpolation algorithm. The method was evaluated using a prototype hemispherical camera designed to mimic the optical properties of an insect compound eye. A dedicated optical flow gantry was used to generate precise three-dimensional trajectories for comparisons to the motion estimated by the hemispherical sensor. The sensor exhibited correct decoupling of translational and rotational motion in four axes, with a local position error of 150mm over an average trajectory length of 1.5m.

1 Introduction

The estimation of egomotion using visual information alone has been a major area of research over the past decades. Many computational techniques have been developed that allow apparent image motion in a plane to be calculated, as well as techniques to derive the motion of the observer given such a pattern of image motion. These techniques are mostly implemented using powerful conventional computers, to allow accurate real-time estimation of egomotion using high resolution sensors. Egomotion from image motion is highly useful in the field of simultaneous localisation and mapping, since it allows the same hardware to be used for both place recognition and motion determination [Nister et al., 2006].

Research into the compound eye of insects has shown that due to its spherical nature, it can provide insects with an almost completely omnidirectional image [Thakoor et al., 2002]. Despite having low-resolution fixed focus vision, insects are capable of regulating their flight speed and position with respect to objects [Srinivasan et al., 1991], estimating the correct time to deploy their legs before landing [Wagner, 1982], and estimating the distance flown using accumulated optical flow [Srinivasan et al., 1996]. It has been proposed that an insect's larger field of view allows for more accuracy in determining total motion, since translational and rotational components are more easily separable [Gluckman et al., 1998].

Inspired by the performance of the spherical insect eye, we have developed a new method to calculate egomotion based on images captured on a spherical image plane. The method derives from the image interpolation algorithm [Srinivasan, 1994], which can be used to compute 3 DoF self-motion from images on a planar camera. We illustrate the extension of the method to 6 DoF, and a solution to the difficulty of non-uniform sensor spacing on a spherical surface, a problem not addressed in previous omnidirectional egomotion estimation research [Vassallo et al., 2002].

The method is evaluated using a unique camera that captures images without a lens or mirror directly on to a hemispherical surface [Maddern et al., 2008]. A dedicated optical flow gantry moved the hemispherical camera through controlled trajectories in four degrees of freedom, in order to evaluate the visual odometry estimation of the sensor. Successful decoupling of motion about the four degrees of freedom was demonstrated, and the perceived trajectories generated by path integration illustrated the response of the sensor to certain environmental conditions, such as proximity to objects.

The paper proceeds as follows. Section 2 briefly describes the methods of finding projected point velocities on a spherical surface, as well as an efficient method of calculating image motion about arbitrary axes on a 2D image plane. Section 3 presents a combination of the above techniques to calculate apparent image motion on a spherical image surface in all 6 degrees of freedom, as well as a method to interpolate an image between irregularly spaced points on a sphere. Section 4 presents the bio-inspired hemispherical image sensor used to evaluate the egomotion algorithms as well as the experimental method used to generate 4DoF trajectories. Results of the visual odometry experiments are presented in Section 5 before the paper concludes in Section 6.

2 Background

In order to determine egomotion using visual data alone, the apparent motion of the image must be calculated. This motion data will produce a vector field of image velocity at every point in the image, known as the optical flow. If the assumption is made that the image generated is the projection of a stationary environment, then all image motion will be due to the motion of the observer. In this case, by using the pattern generated by the velocity vector field the egomotion can be estimated.

2.1 Optical Flow on the Sphere

As presented in [Corke et al., 2009], given a spherical image surface of radius *r*, a world point **P** with image plane co-ordinates of ${}^{c}\mathbf{P} = \{X, Y, Z\}$, the projected co-ordinates *p* on the image surface can be found as follows:

$$\frac{x}{r} = \frac{X}{R}, \frac{y}{r} = \frac{Y}{R}, \frac{z}{r} = \frac{Z}{R} \tag{1}$$

where *R* is the radial distance to the point **P** and *x*, *y* and *z* lie on the image plane such that $\sqrt{x^2 + y^2 + z^2} = r$. If the image plane has translational velocity $v = (v_x, v_y, v_z)$ and angular velocity $\omega = (\omega_x, \omega_y, \omega_z)$, the velocity of any projected point p on the image plane is as follows:

$$\begin{aligned} \dot{x} &= r(z\omega_y - y\omega_z + v_x) \\ \dot{y} &= r(x\omega_z - z\omega_x + v_y) \\ \dot{z} &= r(y\omega_x - x\omega_y + v_z) \end{aligned} \tag{2}$$

By converting the co-ordinates from Cartesian to spherical polar with azimuth angle φ and colatitude angle θ such that:

$$x = z \tan\theta \cos\varphi,\ y = z \tan\theta \sin\varphi \tag{3}$$

The polar form image Jacobian is of the following form:

$$\begin{bmatrix} \dot{\theta} \\ \dot{\varphi} \end{bmatrix} = \frac{1}{Z}\mathbf{J}_t(\theta,\varphi)\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} + \mathbf{J}_w(\theta,\varphi)\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \tag{4}$$

where $\mathbf{J}_t$ and $\mathbf{J}_w$ are as follows:

$$\mathbf{J}_t(\theta,\varphi) = \begin{bmatrix} \cos\varphi\cos^2\theta & \sin\varphi\cos^2\theta & -\cos\theta\sin\theta \\ -\dfrac{\sin\varphi\cos\theta}{\sin\theta} & \dfrac{\cos\varphi\cos\theta}{\sin\theta} & 0 \end{bmatrix} \tag{5}$$

$$\mathbf{J}_t(\theta,\varphi) = \begin{bmatrix} -\sin\varphi & \cos\varphi & 0 \\ \dfrac{\cos\varphi\cos\theta}{\sin\theta} & -\dfrac{\sin\varphi\cos\theta}{\sin\theta} & 1 \end{bmatrix} \tag{6}$$

The presence of the 1/*Z* term in (4) illustrates the limitation of projecting all world points onto a spherical image plane; without explicit knowledge of the distance to points the translational velocity cannot be correctly calculated. However, there is no such constraint on rotational velocity, which can be calculated irrespective of point distance [Corke et al., 2009].

2.2 Image Interpolation Algorithm on a Plane

The image interpolation algorithm [Srinivasan, 1994] generates reference images corresponding to small displacements (translation and rotation around chosen axes), then finds the optimal combination of these reference images that would correspond to the change between two successive frames. This is a non-iterative method that directly calculates the total image motion, and from this the egomotion can be directly estimated.

Given a 2D frame f_0 windowed by a 2D Gaussian Ψ and three reference displacements (Δx_{ref}, Δy_{ref} and $\Delta\theta_{\text{ref}}$), six reference frames can be generated by shifting f_0 in the positive and negative direction of each of those displacements (f_{x+}, f_{x-} etc). By assuming that every successive frame *f* is made up by the previous frame f_0 plus a linear combination of the six reference frames (and thus constraining image motion to the axes of the three reference displacements), the inter-frame displacements Δx, Δy and $\Delta\theta$ can be calculated using the following equations:

$$\begin{aligned} f' = f_0 &+ 0.5\left(\frac{\Delta x}{\Delta x_{\text{ref}}}\right)(f_{x+} - f_{x-}) \\ &+ 0.5\left(\frac{\Delta y}{\Delta y_{\text{ref}}}\right)(f_{y+} - f_{y-}) \\ &+ 0.5\left(\frac{\Delta\theta}{\Delta\theta_{\text{ref}}}\right)(f_{\theta+} - f_{\theta-}) \end{aligned} \tag{7}$$

$$E = \int\int \Psi[f - f']^2\, dx\, dy \tag{8}$$

E is equal to the variance (or mean-square error) between the actual image and the image interpolated from the previous frame, and by finding the partial derivatives of *E* with respect to Δx, Δy and $\Delta\theta$ and setting them equal to zero (corresponding to the error minima), three linear simultaneous equations in three unknowns are formed. By inverting a 3 x 3 matrix containing the coefficients of the three displacements, the changes in *x*, *y* and θ can be estimated.

Additional motion parameters can be calculated by adding additional types of reference image. For example, pitch and yaw rotations could be added by calculating reference images for given rotations around those axes, and image expansion or contraction could be used to calculate *z*-axis translations.

Although this seems like a robust method of calculating total image motion, it is not without constraints. The reference displacements must be small compared to the largest spatial wavelength, or else spatial aliasing will occur. Also, certain types of image motion can yield a singular matrix (such as a featureless or one-dimensional image), however this can be avoided by checking the difference between two reference images in a given axis. Finally, this algorithm cannot determine translation or rotation about an arbitrary axis – all axes must be defined before the interpolation algorithm is applied.

3 Spherical Egomotion Estimation

The algorithm for image interpolation in a plane was intended for use with images generated by a conventional camera, and only defined three degrees of freedom – translation along *x* and *y* and rotation about the *z*-axis. However, three additional motion parameters are required for full 6 DoF motion, consisting of translation along *z* and rotation around the *x*- and *y*-axes. Additionally, a method of generating reference images on a spherical surface is required (along with an interpolation method to map reference image intensities back to the original sensor locations), as well as a method of solving for the six motion variable estimates to minimise the error between sequential frames.

3.1 Image Interpolation on a Sphere

Given a linear displacement Δx, Δy and Δz along the three axes, as well as an angular displacement $\Delta\varphi$, $\Delta\theta$ and $\Delta\psi$ about each axis, the image interpolation algorithm can be

extended to 6DoF as follows:

$$2f' = 2f_0 + \left(\frac{\Delta x}{\Delta x_{\text{ref}}}\right) f_x + \left(\frac{\Delta y}{\Delta y_{\text{ref}}}\right) f_y + \left(\frac{\Delta z}{\Delta z_{\text{ref}}}\right) f_z + \left(\frac{\Delta \varphi}{\Delta \varphi_{\text{ref}}}\right) f_\varphi + \left(\frac{\Delta \theta}{\Delta \theta_{\text{ref}}}\right) f_\theta + \left(\frac{\Delta \psi}{\Delta \psi_{\text{ref}}}\right) f_\psi \quad (9)$$

where f_x, f_y etc. are the difference between the positively and negatively displaced reference images (i.e. $f_x = f_{x+} - f_{x-}$). Since the surface of the spherical camera is continuous, there will be no edge effects and therefore the Gaussian windowing function is not required. Hence, the error between the interpolated frame and the actual frame is given by the following equation:

$$E = \sum_n (f[n] - f'[n])^2 \quad (10)$$

where n is the number of image points or 'photosites' on the surface of the sensor.

3.2 Reference Image Generation

Unlike previous implementations in [Chahl et al., 1996], reference images for the image interpolation algorithm are not directly available. In order to generate a reference image for a given displacement, the field of view of the sensor must be transformed to represent the same image viewed from a different direction. Generation of a reference image for a given rotation is performed by uniformly rotating every image point on the plane, and for a given translation every image point is linearly shifted off the surface of the sphere and then remapped to the image plane by normalising the magnitude of the vector p_i', as shown in Figure 1. Rotation reference images are generated as follows:

$$p_i' = \text{Rot}_{\Delta\theta} p_i \quad (11)$$

Translation reference images are generated as follows:

$$p_i' = \frac{\text{Trans}_{\Delta x} p_i}{|\text{Trans}_{\Delta x} p_i|} |p_i| \quad (12)$$

The reference images produced by applying these transforms for all 6 degrees of freedom are shown in Figure 2. The choice of displacement constants Δx, Δy and Δz must be performed in accordance with the maximum expected translational velocity, the sensor frame rate and an object distance estimate. With a spherical image plane diameter of 50 mm, a 10 mm reference displacement provides velocity estimation of up to 2 m/s for objects 1m from the sensor at a 10 Hz frame rate before spatial aliasing occurs.

3.3 Non-uniform Spacing of Photosites

Given the mechanism for generating virtual fields of view from a number of different reference displacements, a method for directly comparing the two hemispherical images is required. The method outlined in [Srinivasan, 1994] allows direct comparison of reference displacements and actual frames, since they both occupy the same photosites on a 2D plane. However, since the reference

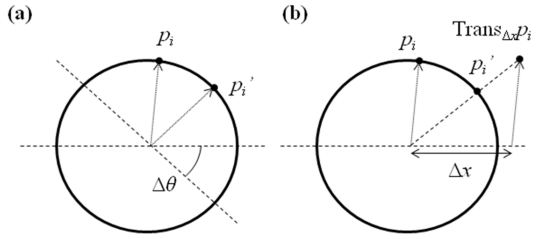

Figure 1 – The location of the point p_i' on the sphere is generated by applying either a rotation transform for image rotation as in (a), or a vector translation and remapping to the image surface for image translation as in (b).

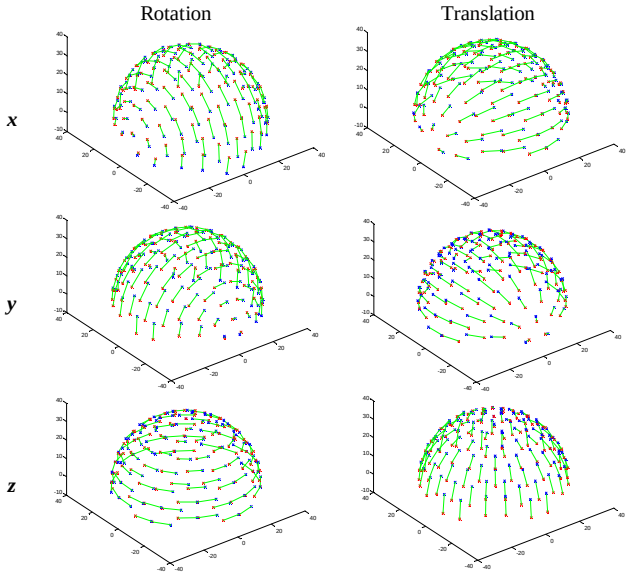

Figure 2 – Reference rotation and translation displacements for a hemispherical imaging plane. Note the similarities between rotation in x and translation in y (and conversely for y rotation and x translation) for the hemispherical case.

images for the spherical image plane contain 'virtual' photosites generated by vector transforms, these will rarely coincide with the original photosites due to their non-uniform spacing on a spherical surface [Krishnan et al., 2009]. A method of interpolating between photosites to determine the expected intensities is needed.

Given a continuous image intensity surface I on the spherical image plane with n discrete intensity measurements at locations $p_{1\ldots n}$, the intensity at any arbitrary point p can be interpolated as follows:

$$I(p) = \frac{\sum_{i=1}^{n} \frac{I(p_i)}{d(p, p_i)}}{\sum_{i=1}^{n} \frac{1}{d(p, p_i)}} \quad (13)$$

where $d(p,p_i)$ is the arc length along the surface of the sphere between points p and p_i, and is found by

$$d(p, p_i) = |p| \text{acos}\left(\frac{p \cdot p_i}{|p|^2}\right) \quad (14)$$

This is derived from the inverse distance weighting proposed in [Shepard, 1968], suited to interpolating between irregularly spaced data points in *n* dimensions.

4 Design of a Hemispherical Camera

The egomotion estimation methods developed above were tested on a hemispherical camera prototype designed to mimic the physiological properties of an insect compound eye, pictured in Figure 3. Each of the 116 phototransistors on the sensor surface has a narrow-angle glass lens and a wide spectral sensitivity, approximating the optical characteristics of compound eye ommatidia. An acrylic outer shell maintains alignment of all sensors such that their optical axes intersect in the centre of the hemisphere, and an onboard microcontroller samples the full image at up to 600 fps, transferring the data to a PC over a high speed USB connection. Further construction details are found in [Maddern et al., 2008].

Figure 3 – Hemispherical image sensor based on the optical characteristics of the insect compound eye. The sensor consists of 116 individually lensed phototransistors arranged on a polyhedral printed circuit board, with outer alignment shell.

Figure 4 shows an image captured by the sensor alongside an omnidirectional image of the same location captured using a conventional camera and a panoramic mirror. While the hemispherical sensor has very low resolution, the lighting and structure of the room are discernible. However, the sensitivities of the individual phototransistors are only approximately calibrated and the image quality would benefit from a more exact intensity calibration method.

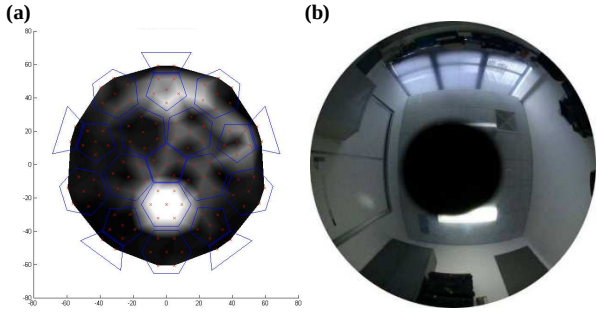

Figure 4 – Comparison between images generated by (a) the hemispherical sensor and (b) conventional omnidirectional image sensor. The occlusion in the photograph is due to the camera lens; the hemispherical sensor has a clear view of the ceiling.

5 Experimental Setup

The 4DoF optical flow gantry pictured in Figure 5 was used to evaluate the performance of the egomotion

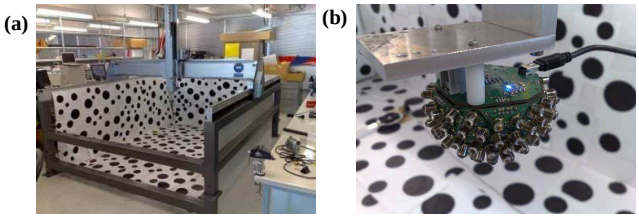

Figure 5 – Optical flow gantry. (a) shows the test pattern (with fourth wall removed) and 4DoF actuation mechanism, and (b) details the sensor attachment. The *z*-axes of the gantry and the sensor intersect to ensure that all motion is separable.

estimation from data captured by the hemispherical sensor. The space enclosed by the walls of the gantry is approximately 2.5 x 2 x 1.5m, and the walls are non-uniformly textured with black circles of between 100 and 200mm diameter.

Eleven different trajectories were used to evaluate the egomotion estimation, with varying combinations of translational and rotational velocities. Reference displacements were set to 10 mm of translation and 10 degrees of rotation. Images were captured at 5 mm intervals in the direction of motion, corresponding to an equivalent forward speed of 3 m/s at the full frame rate of 600 fps.

6 Preliminary Results

The following section presents the egomotion performance statistics for all the trajectories attempted, as well as a detailed analysis of a particular S-bend trajectory. Egomotion in camera-centered co-ordinates and path integration in world co-ordinates are examined separately. As a hemispherical sensor (as opposed to a full spherical sensor) cannot correctly decouple translation and rotation about all six axes, a total of four axes were used (translation along *x*, *y* and *z* and rotation about *z*). A linear scaling factor was applied to the translational motion estimates to provide correct displacements assuming 1 meter distances to all objects.

6.1 Egomotion Estimation

Figure 6 shows camera-centered egomotion estimates for a typical S-bend trajectory. Over the 2 meter trajectory the RMS velocity error was equal to 6.15 mm/frame and 2.29 degrees/frame, and the RMS position error equal to 159.6 mm and 50.4 degrees. While little direct correlation exists in the velocity plots (other than an approximate correlation in *z*-axis rotation at 100 frames) the position tracking over the full trajectory is evident, with the final estimated position within 100 mm and 5 degrees of the final location. Table 1 lists the error estimates for all 11 trajectories attempted. Over an average trajectory length of 1.5 m, the RMS position error was approximately 150 mm and 30 degrees.

6.2 Path Integration

Figure 7 shows the estimated trajectory of the sensor in world co-ordinates, generated by performing path integration over all 4 degrees of freedom. Although the path generated by the sensor does somewhat correspond to

that generated by the gantry (evident in the top-down view on the *x-y* plane), noise in both *z*-axis translation and rotation causes significant degradation to the estimated path. It was found that for other trajectories that significantly varied the distance between the sensor and the textured walls of the gantry, the path integration error was significantly higher.

Table 1 – Egomotion Estimation Error

Trajectory Length	Velocity RMS Error	
	Translation (mm/frame)	Rotation (deg/frame)
2.32	5.24	2.536
0.88	3.54	2.534
0.59	3.08	1.303
2.41	5.6	1.851
1.88	6.15	2.295
1.98	11.57	0.707
1.80	5.37	1.927
0.90	4.95	1.159
2.05	15.97	0.686
0.99	4.1	1.626
2.36	12.59	1.356
Average Error	8.19	1.754

7 Discussion

The preliminary results show the potential for using a small-scale high-speed spherical camera for egomotion estimation. In order to more carefully validate the results, we will need to calibrate the camera to create a uniform response to light from each photosite. At this stage, the 116 phototransistors vary as much as an order of magnitude in gain, which is clearly adding to the noise present in the motion estimates derived from the image interpolation algorithm. Nevertheless, the results clearly show that the correct motion is discernible under the noise created by the variable gain of the sensor.

The chief limitation of the sensor is that the perceived motion is dependent on the environment and the distances to obstacles. Therefore, identical paths in different locations may produce entirely different perceived trajectories, and some sort of compensation is required to correct for the error. Due to this, the sensor may not be suitable for conventional odometry applications, as it does not provide a direct replacement for dead-reckoning systems used in mobile robots. This behavior represents a fundamental limit of the visual odometry algorithm (along with the inability to resolve absolute translational motion), and cannot be resolved without an absolute distance reference.

7.1 Future Work

We intend to apply the egomotion estimation algorithms developed in this paper to a full spherical camera that is built using the same principle as the hemispherical camera described in this paper. While there are some manufacturing challenges to be addressed in producing the full sphere, it will be of immense value to see whether the egomotion estimation algorithm can, in fact, resolve all 6 DoF of self motion from a single camera.

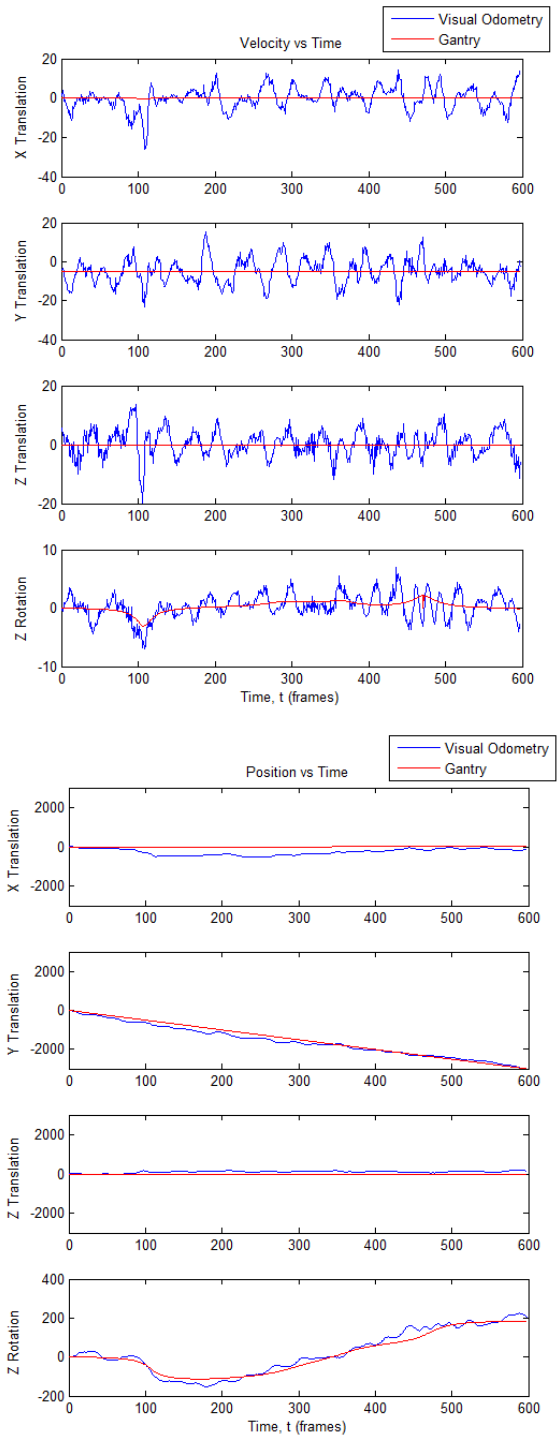

Figure 6 – Visual odometry results for (a) velocity and (b) position in decoupled camera-centred axes. Although the velocity signal is noisy, the position is correctly tracked by the sensor.

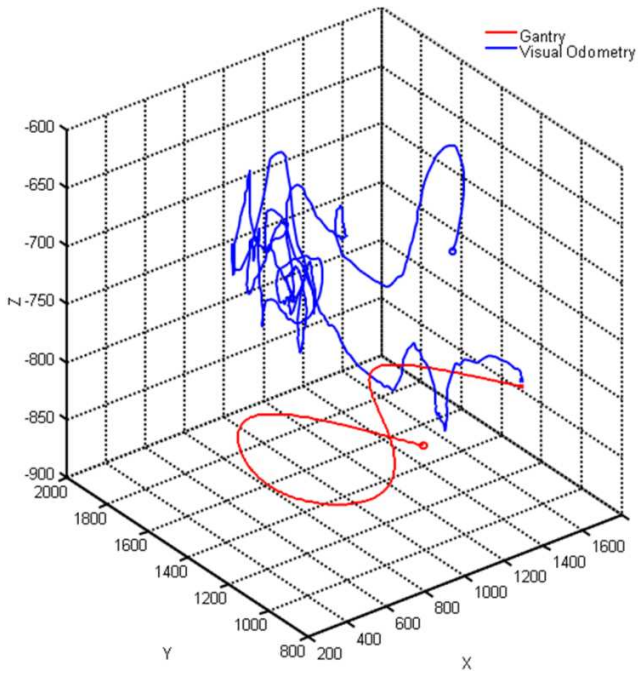

Figure 7 – Path integration performed by the sensor over a gantry trajectory. The path in the *x*-*y* plane shows the sensor follows a roughly correct trajectory, but is coupled with significant *z*-axis position error (amplified here by a change in *z*-axis scale).

Acknowledgment

Will Maddern thanks Professor Mandyam V. Srinivasan for the use of his optical flow gantry to generate the results for this paper and for his advice on various aspects of biologically-inspired visual odometry.

References

[Chahl et al., 1996], J. Chahl and M.V. Srinivasan (1996) Visual computation of egomotion using an image interpolation technique. Biol. Cybernetics 74, 405-411.

[Corke et al., 2009] P. Corke and R. Mahony, "Sensing and Control on the Sphere", in *Proceedings of the 14th International Symposium of Robotics Research*, Lucerne, Switzerland, 31 August - 3 September 2009.

[Gluckman et al., 1998] J. Gluckman and S. K. Nayar, "Ego-motion and omnidirectional cameras," in *Sixth International Conference on Computer Vision*, 1998, pp. 999-1005.

[Krishnan et al., 2009] G. Krishnan and S. Nayar, "Towards a true spherical camera," *Human Vision and Electronic Imaging XIV*, G. Krishnanand S. Nayar (eds), *Proceedings of the SPIE*, volume 7240, pp. 724002-724002-13, 2009.

[Maddern et al., 2008] W. Maddern and G. Wyeth, "Development of a Hemispherical Compound Eye for Egomotion Estimation," in *Proceedings of the 2008 Australian Conference on Robotics and Automation*, Canberra, Australia, December 2008.

[Nister et al., 2006] D. Nister, O. Naroditsky and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol 23, issue 1, pp. 3-20, Jan 2006.

[Shepard, 1968] Shepard, "A two-dimensional interpolation function for irregularly-spaced data," *in Proceedings of the 23rd Association for Computing Machinery National Conference*, 1968, pp. 517-524.

[Thakoor et al., 2002] S. Thakoor, J. Chahl, M. V. Srinivasan, L. Young, F. Werblin and B. Hine, "Bioinspired engineering of exploration systems for NASA and DoD," *Artificial Life*, vol 8, issue 4, pp. 357-369, 2002.

[Srinivasan, 1994] M. V. Srinivasan, "An image-interpolation technique for the computation of optic flow and egomotion," *Biological Cybernetics*, vol 71, pp. 401-415, 1994.

[Srinivasan et al., 1991] M. V. Srinivasan, M. Lehrer, W. H. Kirchner and S. W. Zhang, "Range perception through apparent image speed in freely-flying honeybees," *Visual Neuroscience*, vol 6, pp. 519-535, 1991.

[Srinivasan et al., 1996] M. V. Srinivasan, S. W. Zhang, M. Lehrer and T. S. Collett, "Honeybee navigation en route to the goal: visual flight control and odometry," *Journal of Experimental Biology*, vol 199, pp. 237-244, 1996.

[Vassallo et al., 2002] Vassallo, R.F., Santos-Victor, J., Schneebeli, H.J., "A general approach for egomotion estimation with omnidirectional images," Third Workshop on Omnidirectional Vision, 2002, pp. 97--103.

[Wagner, 1982] H. Wagner, "Flow-field variables trigger landing in flies," *Nature* (London), vol 297, pp. 147-148, 1982.