

Robust Ego-Motion Estimation

D.A.Sinclair, A.Blake, D.W.Murray

Robotics Research Group
Department of Engineering Science
Oxford University
Oxford OX1 3PJ

This paper aims to develop a simple algorithm for determining the direction of ego-motion of a monocular observer in a static environment. Rather than returning a single direction the algorithm returns a set of motion directions known to contain the true direction. Translational motion can be arbitrary. The algorithm is based on one invented by Horn and Weldon [1]. It uses only normal components of optical flow at discrete image locations and is only intended to function when the contribution of rotation to the apparent motion is small. The algorithm uses the fact that imaged points must be at some positive distance from the viewer. No implicit assumptions are made about the scene being viewed. The algorithm is illustrated on an artificially generated set of normal components. It is then tested on a real image sequence. By comparison with [1], this paper develops the idea of introducing rotation tolerance at the expense of accuracy of ego-motion determination. An analysis of the shape of the resulting set of possible motion directions is given.

In this paper the problem of determining the motion direction of a monocular observer with respect to a rigid, unknown world is considered. Ordinarily the full optical flow at several image locations is required for motion estimation, the aperture problem renders the recovery of full optical flow difficult. The standard techniques of velocity smoothing or planar approximation have obvious weaknesses. This paper shows that only the normal components of optical flow, (or equivalently first order spatio-temporal derivatives of image brightness), are required for motion direction estimation. The algorithm developed by Horn and Weldon is valid only when rotational velocity is known. This paper provides a method of making Horn's algorithm robust to small unknown rotations at the expense of the accuracy of the motion estimate. The algorithm returns a set of possible motion directions known to contain the true motion direction, rather than a single estimate. For simplicity the paper concentrates on the process of estimating viewer motion direction from a set of normal components of flow, rather than the process of generating these components. The

proposed algorithm is first demonstrated on synthetic data to illustrate the effects of noise and small rotations, and subsequently tested on real data generated from images taken with a moving camera. One possible motion direction estimator is presented. The method also suggests an alternative method for finding the intersection of the optical axis and the imaging array of the camera.

CONVERSION TO SPHERICAL PROJECTION

Spherical rather than planar projection greatly simplifies the mathematics of this algorithm. Flow vectors generated from images captured in planar projection are readily transformed to their equivalents in spherical projection. If \mathbf{r} is the co-ordinate of a point in the world it will be projected to $\mathbf{P} = \mathbf{r}/Z$ in planar projection and to $\mathbf{Q} = \mathbf{r}/r$ in spherical projection then:

$$\dot{\mathbf{Q}} = \mathbf{P} \cdot \mathbf{Q}(\dot{\mathbf{P}} - (\mathbf{Q} \cdot \dot{\mathbf{P}})\mathbf{Q}). \quad (1)$$

The equation relating the full flow vector \mathbf{V} to the translational velocity \mathbf{U} , the angular velocity Ω of the imager and the distance to points in the world r at image position \mathbf{Q} is [3]:

$$\mathbf{V} = (\mathbf{U} - (\mathbf{Q} \cdot \mathbf{U})\mathbf{Q})/r + \Omega \times \mathbf{Q}, \quad (2)$$

in spherical projection with focal length scaled to unity and with \times denoting the vector product. The equation for the normal component of flow (in spherical projection) is:

$$\mathbf{V} \cdot \mathbf{n} = \mathbf{U} \cdot \mathbf{n}/r + (\mathbf{Q} \times \mathbf{n}) \cdot \Omega, \quad (3)$$

where \mathbf{n} denotes the unit normal to an edge element in the image sphere. *ie* the vectors \mathbf{n} are all perpendicular to their individual image locations \mathbf{Q} . For mathematical convenience the viewer is assumed to have a circular field of view centered on the optical axis of the viewer.

USING THE ($r > 0$) CONDITION

If $|\Omega|$ in (3) is equal to zero or is negligible then:

$$\text{sign}(\mathbf{V} \cdot \mathbf{n}) = \text{sign}(\mathbf{U} \cdot \mathbf{n}/r), \quad (4)$$

since r the distance to imaged points is positive. Equivalently, when Ω is accurately known by letting

$$\mathbf{V} \cdot \mathbf{n} - (\mathbf{Q} \times \mathbf{n}) \cdot \Omega = \mathbf{V}' \cdot \mathbf{n} \quad (5)$$

then,

$$\text{sign}(\mathbf{V}' \cdot \mathbf{n}) = \text{sign}(\mathbf{U} \cdot \mathbf{n}). \quad (6)$$

We define a new vector,

$$\mathbf{N} = [\text{sign}(\mathbf{V}' \cdot \mathbf{n})]\mathbf{n}. \quad (7)$$

There will in general exist a set of vectors \mathbf{N}_i corresponding to the components of flow normal to a set of brightness edges at positions \mathbf{Q}_i in any image taken from a motion sequence. If the small rotation assumption is true then each of the vectors \mathbf{N}_i must satisfy the condition

$$\mathbf{U} \cdot \mathbf{N}_i \geq 0. \quad (8)$$

Therefore each \mathbf{N}_i lies within 90° of \mathbf{U} , (\mathbf{N}_i is then a three dimensional unit vector.) The set of \mathbf{N}_i is confined to one half of a unit-sphere with \mathbf{U} as its pole. Finding the direction of viewer motion corresponds to finding the pole of the hemisphere containing all \mathbf{N}_i , this direction is termed $\hat{\mathbf{U}}$.

THE DOMAIN OF N

The size of the field of view will limit the range of directions that \mathbf{N} may take. If the field of view of the viewer is $2\theta_v$, then any image vector \mathbf{Q} must lie within a cone of directions with half angle θ_v around the viewers optical axis. Any vector \mathbf{N}_i is by definition perpendicular to its associated image position \mathbf{Q}_i . Let

$$\mathbf{t}_i = (\mathbf{Q}_i \times \mathbf{N}_i) \quad (9)$$

hence for any i \mathbf{Q}_i , \mathbf{N}_i and \mathbf{t}_i form an orthogonal triad. If \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z are the unit vectors of the cartesian axes then (from Horn [1]):

$$(\mathbf{Q}_i \cdot \mathbf{e}_z)^2 + (\mathbf{N}_i \cdot \mathbf{e}_z)^2 + (\mathbf{t}_i \cdot \mathbf{e}_z)^2 = 1 \quad (10)$$

and

$$(\mathbf{N}_i \cdot \mathbf{e}_x)^2 + (\mathbf{N}_i \cdot \mathbf{e}_y)^2 + (\mathbf{N}_i \cdot \mathbf{e}_z)^2 = 1. \quad (11)$$

Hence

$$(\mathbf{Q}_i \cdot \mathbf{e}_z)^2 + (\mathbf{t}_i \cdot \mathbf{e}_z)^2 = (\mathbf{N}_i \cdot \mathbf{e}_x)^2 + (\mathbf{N}_i \cdot \mathbf{e}_y)^2. \quad (12)$$

However \mathbf{Q}_i must lie within θ_v of the optical axis \mathbf{e}_z , so that

$$(\mathbf{Q}_i \cdot \mathbf{e}_z)^2 \geq \cos^2(\theta_v) \quad (13)$$

and since $(\mathbf{t}_i \cdot \mathbf{e}_z)^2 > 0$, all vectors \mathbf{N}_i must satisfy

$$(\mathbf{N}_i \cdot \mathbf{e}_x)^2 + (\mathbf{N}_i \cdot \mathbf{e}_y)^2 \geq \cos^2(\theta_v) \quad (14)$$

The vectors \mathbf{N}_i must therefore lie within a 'permissible band' of half width θ_v around the equator of the unit sphere as well as to one side of the plane with \mathbf{U} as its

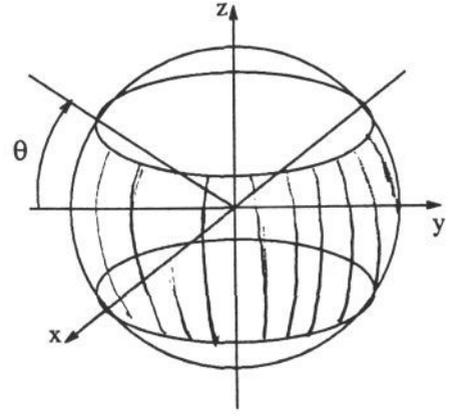


Figure 1: The 'permissible band'. Every unit normal to an edge, in the image, \mathbf{n} is constrained to lie within a permissible band on the unit sphere. For an ideal viewer in spherical projection \mathbf{e}_z , the optical axis of the viewer, is the axis of the permissible band. The band lies between two latitudes $\pm\theta_v$ from the equator, where $\pm\theta_v$ is the field of view of the camera.

normal. This 'permissible band' is shown in figure 1. Figure 2 shows the domain of \mathbf{N} for a given motion direction. The set of vectors \mathbf{N}_i lie on a unit sphere. This set can be visualized by orthographically projecting, (along the x axis), those \mathbf{N}_i with $\mathbf{N}_i \cdot \mathbf{e}_x > 0$ into one view and the others into a second view. The two images in figure 3 are an example of this. The estimated direction of ego-motion $\hat{\mathbf{U}}$, corresponds to the pole of the hemisphere that contains all the vectors \mathbf{N}_i .

THE IMPORTANCE OF A WIDE FIELD OF VIEW

The field of view θ_v , determines the width of the permissible band on the \mathbf{N} sphere. Figures 3 and 4 demonstrate the effect of varying the field of view. The direction of the pole of the hemisphere containing the set of vectors \mathbf{N}_i is determined by \mathbf{U} , the motion direction. If the motion direction (or equivalently the focus of expansion) is contained within the field of view then the boundary of the hemisphere containing all \mathbf{N}_i lies within the permissible band, on the \mathbf{N} sphere. If the only constraint used is the fact that $r > 0$ for each \mathbf{N} , then each \mathbf{N} has equal say in the selection of $\hat{\mathbf{U}}$, the motion estimate. An algorithm is needed which positions a hemisphere so as to encompass the maximum number of \mathbf{N}_i . Points within the hemisphere away from its boundary will not influence its final position, (algorithms must not be vulnerable to clusters of \mathbf{N}_i .) Since only points near the boundary can play a part in positioning the hemisphere it is desirable to have points near to the entire length of the boundary. In the case of a noise and rotation free set of \mathbf{N}_i vectors, a hemisphere with pole in the direction of \mathbf{U} will neatly

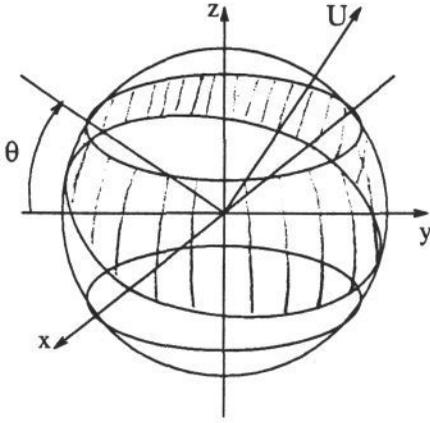


Figure 2: The domain of possible N directions. The vectors N are constrained (by the results of the second section) to lie within a permissible band of width θ_v . They are also constrained to lie on one side of the plain through O with normal parallel to the motion direction U . The region satisfying both conditions is shaded.

encompass the set. This hemisphere will be unique provided that the vectors N_i are uniformly distributed even if the boundary of the hemisphere lies partially outside the permissible band. In the case of noisy data points will tend to 'spill out' from this hemisphere leaving a range of possible hemispheres each of which contain all but a few of the N_i . As in [4] 'slant' and 'tilt' directions are defined to describe the motion direction relative to the optical axis of the viewer. Slant will be taken to refer to the angle between the motion direction and the optical axis and tilt to the azimuthal angle about the optical axis. If the field of view is narrow and U lies outside the field of view then there will exist a large range of possible \hat{U} for which the associated hemisphere contains all but a few of the N_i . The set of possible \hat{U} will have a greater range of slant directions than tilt directions. Reducing the field of view will enlarge the set of possible motion directions in the slant direction. Therefore it is desirable to have a wide field of view to deal with a large range of translation directions.

BUILDING IN TOLERANCE TO ROTATION

The rotation dependent term in equation (3) is $(Q \times n) \cdot \Omega$, because Q and n are orthogonal the maximum value this term can take is $|\Omega|$. The constraint set of N vectors can be reduced by only accepting points where:

$$|V \cdot n| > \Omega_{max}, \quad (15)$$

Ω_{max} being the desired magnitude of rotation to which the algorithm will be tolerant. Removing such N_i is sufficient to ensure that the remaining ones satisfy (8). In the case where rotation is absent the effect of thresholding

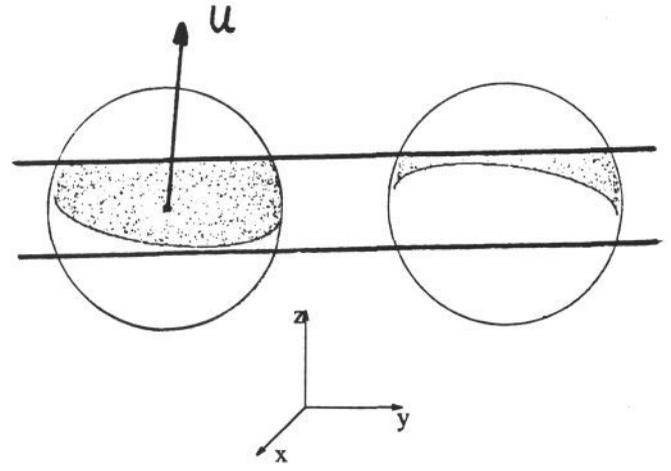


Figure 3: Visualisation of the set of vectors N . The set may be divided into two halves, according to the sign of $e_x \cdot N$. Each half is projected orthographically along the positive or negative x axis. The left image corresponds to those N with $e_x \cdot N > 0$, ie the front half of the unit sphere. The great circle of vectors N satisfying $U \cdot N = 0$ is also projected. This great circle corresponds to the edge of a hemisphere containing all N . The pole of this hemisphere is the motion direction. In this case the simulated motion direction is $U = (0.321, 0.177, 0.940)$ and $\theta_v = 25^\circ$. Estimating motion direction \hat{U} is equivalent to finding the pole of a hemisphere containing all N .

the constraint set is shown in figure 5. The translational flow term $(U \cdot n)/r$ can be thought of as:

$$V \cdot n = |U| \cos(\delta)/r \quad (16)$$

where δ is the angle between U and n . The value of $V \cdot n$ will then tend to be symmetric about U . A thresholding angle δ_T can be defined:

$$\delta_T = \cos^{-1}(r_{min} \Omega_{max}/|U|) \quad (17)$$

Thresholding the set of N vectors according to the value of $V \cdot n$ will remove those N with δ between δ_T and 90 degrees (it will also tend to remove N associated with large r). This thresholded set of vectors no longer uniquely fills one hemisphere, there now exists a set of hemispheres containing all N . Corresponding to this set of hemispheres there exists a set of permissible motion directions. These directions (in the case where the motion direction lies within the field of view) will be contained in a cone of half angle

$$\theta_c = 90 - \delta_T \quad (18)$$

and this cone will be centered on the true motion direction U . However when the boundary of the constraint set lies partially outside the permissible band and the set is thresholded, the set of possible motion directions will correspond to a cone with oblate rather than circular cross-section. Simple analysis shows that, in the worst case of translation perpendicular to the motion direction, the ratio of the widths of the set of \hat{U} , in the slant and

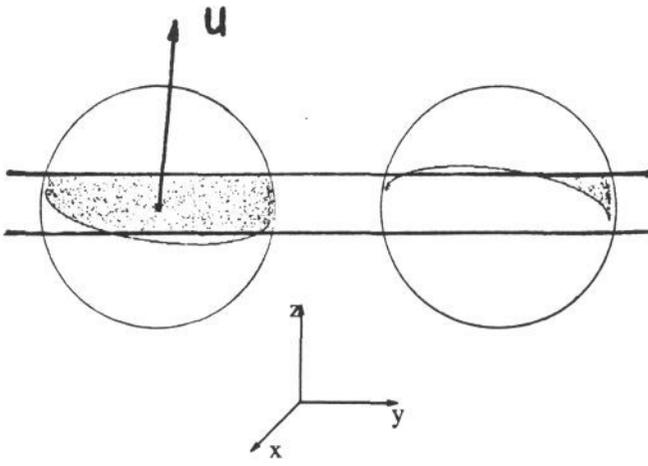


Figure 4: Changing the field of view. If the field of view is reduced so the width of the permissible band is reduced. If the motion direction lies outside the field of view then part of the boundary of the hemisphere containing all \mathbf{N} will lie outside the permissible band. Again $\mathbf{U} = (0.321, 0.177, 0.940)$ but $\theta_v = 15^\circ$. Part of the projected boundary of the hemisphere containing all \mathbf{N} can be seen to lie outside the permissible band.

tilt directions is $1/\theta_v$. Figure 6 illustrates this. In this case decreasing θ_v reduces the accuracy with which the 'slant' direction of motion is determined without altering the accuracy of the tilt estimate. This compares with a result of Maybank in [4] from an entirely different kind of analysis. He showed that when the maximum number of points, five, required for determining ego-motion is reduced to four, the accuracy of the slant of the F.o.E. deteriorates markedly more than tilt.

INTRODUCTION OF ROTATION

Can a Constraint Set Exist

In general the above method will break down when $\text{sign}(\mathbf{V} \cdot \mathbf{n}) = \text{sign}(\mathbf{U} \cdot \mathbf{n})$ is no longer true for the majority of \mathbf{n} . The condition:

$$|\Omega| > |\mathbf{U}|/2r_{min} \quad (19)$$

is sufficient to render this untrue. As Ω increases $|\mathbf{V} \cdot \mathbf{n}|$ ceases to be a single valued function of the vector \mathbf{n} . For any given normal in the image \mathbf{n} there exist an arc of a great circle of possible \mathbf{Q} directions within the image sphere, from which \mathbf{n} could have come. The normal to the plane containing this great circle will be the vector \mathbf{n} . Hence the vector $\mathbf{Q} \times \mathbf{n}$ will be tangent to this arc of a great circle and $(\mathbf{Q} \times \mathbf{n}) \cdot \Omega$ (in equation (3)) will not be a constant. The maximum variation in $(\mathbf{Q} \times \mathbf{n}) \cdot \Omega$ on any great circle through the center of the field of view is $2|\Omega| \cos \theta_v$.

Rotations Effect on the Center of the Set Possible Motion Directions

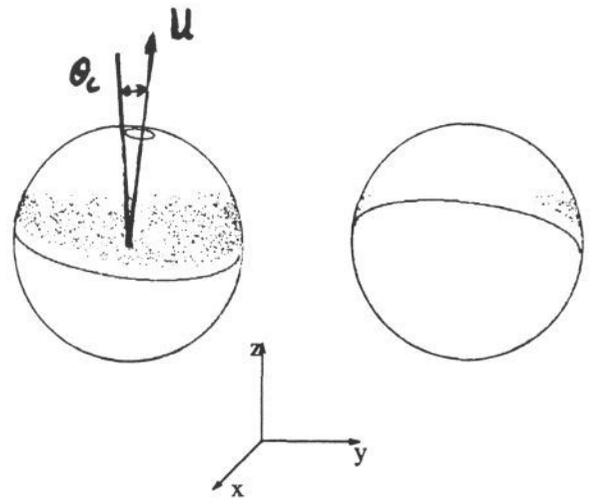


Figure 5: Building in tolerance to rotation. If vectors \mathbf{N} are rejected when $|\mathbf{V} \cdot \mathbf{n}| < |\Omega_{max}|$, then the set no longer fills a hemisphere. There now exists a set of possible motion directions $\hat{\mathbf{U}}$. If the actual motion direction lies inside the field of view this set of motion directions forms a cone. The cone of $\hat{\mathbf{U}}$ found by the author's algorithm, with half angle $\theta_c = 7.6^\circ$ is shown. The cone is centered on $(0.321, 0.177, 0.940)$. the threshold chosen was 20% of $|\mathbf{V} \cdot \mathbf{n}|_{max}$.

The center of the cone of possible translation directions will not correspond to the motion direction \mathbf{U} if a flow field \mathbf{V} , generated by rotation, resembles one resulting from translation. This happens when rotation is present about the viewer x or y axes. The cone of possible $\hat{\mathbf{U}}$ will still contain the translation direction but its center will differ from the \mathbf{U} . As, after thresholding, all the remaining vectors \mathbf{N} again satisfy 8.

Restrictions on Use for Motion Estimation

If a vehicle is to follow a curved path there will be an element of imager angular velocity introduced simply from the curvature of the path;

$$|\Omega| = |\mathbf{U}|/\rho \quad (20)$$

where ρ is the radius of curvature of the path. If

$$\mathbf{U} \cdot \mathbf{n}/r \gg (\mathbf{Q} \times \mathbf{n}) \cdot \Omega \quad (21)$$

is to be true, then

$$\rho/r \gg 1 \quad (22)$$

or the radius of curvature of the vehicle path must be greater than the distance to imaged points.

The Effect of Noise

If noise is present in the data it may be treated in a similar way to rotation. Noise added to $\mathbf{V} \cdot \mathbf{n}$ will tend to

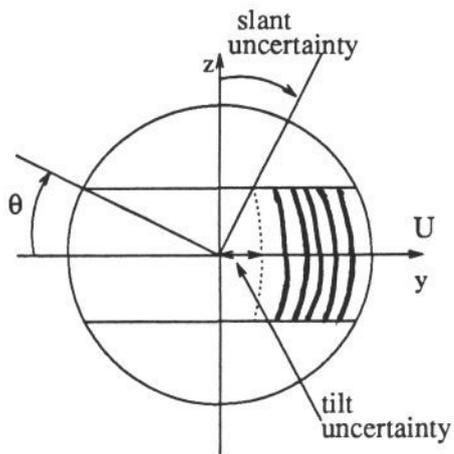


Figure 6: The shape of the set of possible translations directions \hat{U} . If the motion direction lies outside the field of view then, when tolerance to rotation is introduced, the range of permissible slant values is greater than the range of permissible tilt values. For motion perpendicular to e_z the ratio of uncertainty in slant and tilt motion is $1/\theta_v$. The dotted line in the figure represents the boundary of the set of N after thresholding.

spread the constraint set out. Figure 7 shows the effect of adding uniformly distributed noise between $\pm 20\%$ of $|\mathbf{V} \cdot \mathbf{n}|_{max}$ to the simulated data. Figure 8 shows the set restored through thresholding. However the algorithm is less tolerant to Gaussian noise as it tends to generate 'outliers' that remain after some thresholding. A noisy constraint set may still more than fill a hemisphere and hence, if the data is too noisy, looking for cones of possible translation directions is apt to be fruitless.

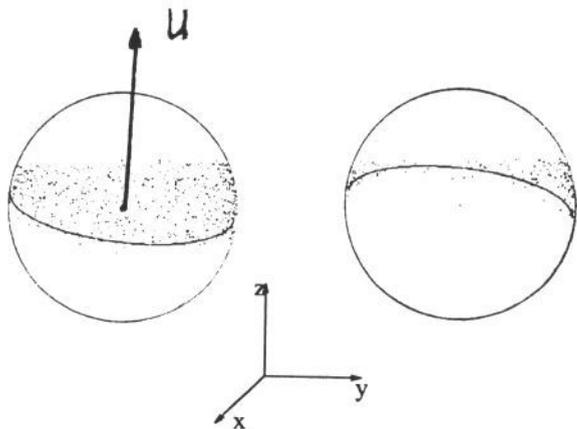


Figure 7: Introduction of noise. Uniformly distributed noise 20% of $\mathbf{V} \cdot \mathbf{n}_{max}$ was added to $\mathbf{V} \cdot \mathbf{n}$. This has caused points to spill out from the hemisphere. Again $\mathbf{U} = (0.321, 0.177, 0.940)$ and the curve in the image represents the edge of a hemisphere with pole parallel to \mathbf{U} .

RESULTS ON REAL IMAGERY

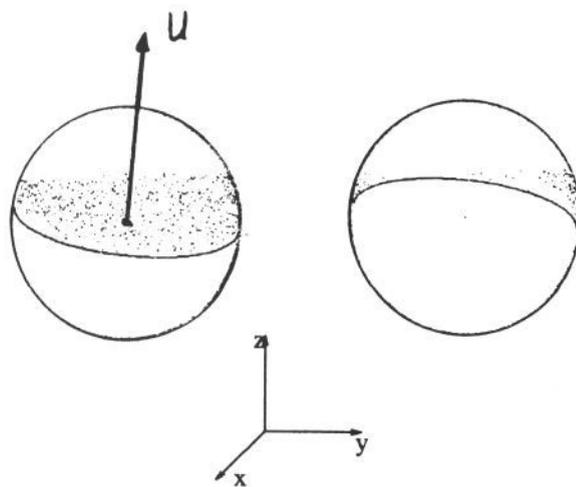


Figure 8: Restoring noisy data. The noisy set of N shown in figure 7 is thresholded at 20% of $\mathbf{V} \cdot \mathbf{n}_{max}$. All the vectors N now lie within one hemisphere again.

The author was lucky enough to have access to a set of vernier velocities of Canny edges generated by a camera in accurately known linear motion. The scene featured in the motion sequence is shown in figure 9. The projection of the resulting unthresholded set of N vectors is shown in figure 10, together with the real and recovered motion direction. The thresholded set is also shown in figure 11 together with the associated cone of possible motion directions.



Figure 9: Scene of a laboratory. The image shown is taken from the motion sequence used to generate the set of N in figures 10 and 11. The camera used was mounted on an optical bench and translated parallel to its own optical axis towards the scene.

USING THE PERMISSIBLE BAND FOR CAMERA CALIBRATION

The orientation of the permissible band within the unit N sphere is determined by the position of the axis of the detector relative to the optical axis of the viewer.

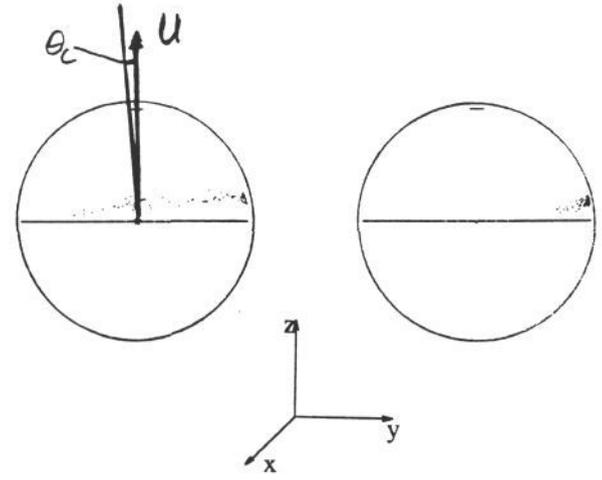
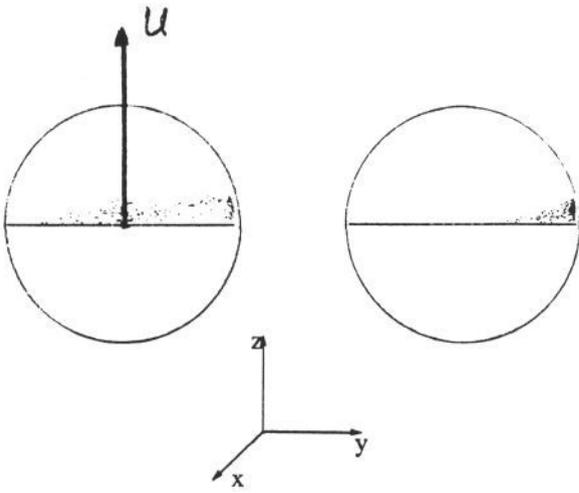


Figure 10: *Real data. The set of \mathbf{N} vectors displayed is unthresholded. The vectors were obtained by matching edge elements between images. All \mathbf{N} fell within the hemisphere with $(0, 0, 1)$ as its pole. The true motion direction was $\mathbf{U} = (0, 0, 1)$. The field of view $\theta_v = 18^\circ$.*

Figure 11: *The set of \mathbf{N} is thresholded at 14% of $\mathbf{V} \cdot \mathbf{n}_{max}$. A cone of possible motion directions with half angle $\theta_c = 3.5^\circ$, centered on $(0.012, 0.012, 1.000)$ was recovered. This set does contain the true motion direction. For $|\mathbf{U}| = 1\text{m.s}^{-1}$ and $r = 1\text{m}$ the algorithm is immune to rotations up to 2deg.s^{-1} . The corresponding minimum radius of curvature of any path ρ is then 25m .*

If the optical axis is taken to be the z axis of the \mathbf{N} sphere and z' taken to be the axis of the detector then new basis vectors x' and y' may be defined. Identical analysis to second section can be performed to show that the permissible band now has z' as its axis. If motion is known relative to the optical axis then

$$\mathbf{e}'_z = \mathbf{e}_z + \hat{\mathbf{U}} - \mathbf{U} \quad (23)$$

will be the vector through the center of the detector and the optical center of the viewer.

ACKNOWLEDGMENTS

Members of staff of the Oxford University Robotics Research Group in particular Prof. J. M. Brady, R. Cipolla and Dr A. Zisserman are gratefully acknowledged for their enthusiasm and advice.

SUMMARY

The paper demonstrates the feasibility recovering a set of directions known to contain the ego-motion direction, solely from the normal components of optical flow, even in the presence of rotation.

References

- [1] B.K.P.Horn, J.E.Weldon Jr 1987, Computationally Efficient Methods for Recovering Translational Motion, IEEE ICCV.
- [2] B.K.P.Horn 1986, Robot Vision, MIT Press.

- [3] S.J.Maybank 1985, The Angular Velocity Associated the Optical Flow Field from Motion Through a Rigid Environment, Proc. Roy. Soc. pp 317 - 326.
- [4] S.J.Maybank 1987, A Theoretical Study of Optic Flow, PhD Thesis, Birkbeck College, University of London.
- [5] E.C.Hildreth 1984, The Measurement of Visual Motion MIT Press. Cambridge, Mass.
- [6] B.F.Buxton, H.Buxton and D.W.Murray 1984, 3D solutions to the aperture problem, Proceedings of the 6th European Conference on Artificial Intelligence, pp 631 - 640.
- [7] H.G.Longuet-Higgins 1980, The Interpretation of a Moving Retinal Image Proc. Roy. Soc. pp 385 - 397.