Relating Egomotion and Image Evolution

Barak A. Pearlmutter

Leonid Gurvits*

March 1995
Report Number SCR-95-TR-536

Abstract By considering the dynamics of the apparent motion of a stationary object relative to a moving observer, we construct a partial differential equation that relates the changes in an image to the motion of the observer. These come in two varieties: a first order system that describes the coevolution of the egocentric radial distances to objects and the visual scene, and a second order system that does not involve any distances or other geometry. The later equation leads, *via* the calculus of variations, to a novel technique for recovering egomotion from image sequences, a so-called visual yaw detector, which is tested on real data. For expository purposes the derivation is carried out in two dimensions, but the approach extends immediately to three.

*Present address: NEC Research Institute, 4 Independence Way, Princeton, NJ 08540

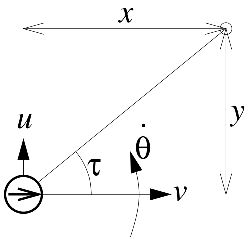
1. Introduction

Using a special camera mounted on the roof of a motorcar which gives a narrow 360 degree strip along the horizon, we are interested in recovering the angular and forward velocity of the vehicle. We sometimes know the velocity from other sensory modalities, such as a sensor attached to the speedometer or odometer. In that case, we need recover only the angular velocity. To this end, we have developed a theory of the evolution of such images. Because of the particular domain of interest, we are primarily interested in the two-dimensional world described above. However, the techniques we develop extend immediately to three dimensions.

For a full description of the visual sensor based navigation system, see Gorr *et al.* (1995), Novak *et al.* (1995), Novak and Hancock (1995), Lin and Judd (1995) or pending patents (Siemens Docket No.s 93E7601, 94E7541, 94E7617, 94E7618 US.)

2. Point movement

Consider an observer in a 2D space moving forward with velocity v, leftward with velocity u, and turning counterclockwise with rotational velocity $\dot{\theta}$. This observer sees a fixed point at angle τ counterclockwise from the forward direction. We wish to relate v, u, $\dot{\theta}$ and τ.



If the fixed point is x ahead of the observer and y to the side, then by definition $\tan\tau = y/x$. By the linearity of differentiation $dx/dt = \dot{\theta}y - v$ and $dy/dt = -\dot{\theta}x - u$. Taking the derivative with respect to time of $y\cos\tau = x\sin\tau$ gives

$$v\sin\tau - u\cos\tau = (\dot{\tau} + \dot{\theta})(x\cos\tau + y\sin\tau) \tag{1}$$

and taking the derivative with respect to time again gets rid of x and y to yield

$$\dot{v}\sin\tau - \dot{u}\cos\tau = \frac{\ddot{\theta} + \ddot{\tau}}{\dot{\theta} + \dot{\tau}}(v\sin\tau - u\cos\tau) - (\dot{\theta} + 2\dot{\tau})(v\cos\tau + u\sin\tau) \tag{2}$$

In the special case of no lateral motion $u = 0$ so this simplifies to

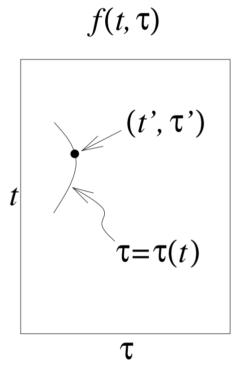
$$\ddot{\theta} + \ddot{\tau} = (\dot{\theta} + \dot{\tau})(\dot{v}/v + (\dot{\theta} + 2\dot{\tau})\cot\tau) \tag{3}$$

which can also be expressed as the more mnemonic

$$\frac{d}{dt}\log\frac{v}{\dot{\theta} + \dot{\tau}} + (\dot{\theta} + 2\dot{\tau})\cot\tau = 0 \tag{4}$$

3. Isointensity contours

We would like to characterize the contour lines of a function $f(t, \tau)$, so that we can check if they satisfy some given differential equation.



Consider an arbitrary point (t', τ'). For notational convenience, let $\tau(t)$ be the contour line that runs through (t', τ'), so for instance $\tau' = \tau(t')$. Since $\tau(t)$ is a contour line, by definition $f(t, \tau(t))$ is constant. Using subscripts to denote partial derivatives, $\frac{d}{dt}f(t, \tau(t)) = f_t + f_\tau\dot{\tau}(t) = 0$ so

$$\dot{\tau} = -f_t/f_\tau \tag{5}$$

is the slope of the contour line through an arbitrary point (t, τ).

Similarly, taking the derivative with respect to t again,

$$\frac{d^2}{dt^2}f(t, \tau(t)) = f_{tt} + \dot{\tau}f_{t\tau} + f_\tau\ddot{\tau} + (f_{t\tau} + f_{\tau\tau}\dot{\tau})\dot{\tau} = 0$$

so

$$\ddot{\tau} = -f_{tt}/f_\tau + 2f_tf_{t\tau}/f_\tau^2 - f_t^2f_{\tau\tau}/f_\tau^3 \tag{6}$$

gives the second derivative of the contour line.

3.1. Image evolution

Let $f(t, \tau)$ be the image intensity measured at time t angle τ. Assuming Lambertian reflection and no occlusion, the contour lines of f are the angular paths taken by particular visible points in the world. Hence the contour line going through each point

(t, τ) must satisfy equation 3. We therefore substitute in the first two derivatives of the contour line. This results in a condition on the partial derivatives of f,

$$\ddot{\theta} f_\tau^3 - (f_{tt} f_\tau^2 - 2 f_t f_{t\tau} f_\tau + f_t^2 f_{\tau\tau}) = (\dot{\theta} f_\tau - f_t) f_\tau (f_\tau \dot{v}/v + (\dot{\theta} f_\tau - 2 f_t) \cot \tau) \tag{7}$$

This equation will be obeyed wherever the assumptions made above are valid.

4. Distances

4.1. Distance from local gradients

The distances to the observed object in direction τ at time t has surprisingly simple form. Substituting $R = x \cos \tau + y \sin \tau$ into equation 1 and solving for R gives

$$R = \frac{v \sin \tau - u \cos \tau}{\dot{\theta} + \dot{\tau}} \tag{8}$$

which can be put in terms of image intensity

$$R = f_\tau \frac{v \sin \tau - u \cos \tau}{\dot{\theta} f_\tau - f_t} \tag{9}$$

4.2. First order system for image evolution

From equation 8 we have

$$\dot{\tau} = (v \sin \tau - u \cos \tau)/R - \dot{\theta}.$$

Let $g(t, \tau)$ be some property we measure of a patch of an object at time t viewing angle τ. If we know how this property changes as a function of viewing angle and egomotion, then we know

$$\frac{d}{dt} g(t, \tau(t)) = g_t + \dot{\tau} g_\tau = \dot{g}(\tau)$$

$$g_t + ((v \sin \tau - u \cos \tau)/R - \dot{\theta}) g_\tau = \dot{g}(\tau) \tag{10}$$

where $\tau(t)$ is the viewing angle of the patch under consideration and $\dot{g}(\tau)$ is the temporal derivative of the property of interest.

For image intensity $g = f$, and under the lighting and reflectance assumptions made above $\dot{f} = 0$. Hence

$$f_t = ((u \cos \tau - v \sin \tau)/R + \dot{\theta}) f_\tau$$

More interesting perhaps is when we let $g = R$. We must calculate $\dot{R}$, which is straightforward. $R^2 = x^2 + y^2$ so $R\dot{R} = x\dot{x} + y\dot{y} = x(\dot{\theta} y - v) + y(-\dot{\theta} x - u) = -vx - uy$ and since $x = R \cos \tau$ and $y = R \sin \tau$ we havge $\dot{R} = -v \cos \tau - u \sin \tau$ and therefore

$$R_t + v \cos \tau + u \sin \tau = ((u \cos \tau - v \sin \tau)/R + \dot{\theta}) R_\tau$$

Substituting $r = 1/R$ into this last pair of equations with $R_t = -r_t/r^2$ yields

$$f_t = ((u \cos \tau - v \sin \tau) r + \dot{\theta}) f_\tau \tag{11}$$

$$r_t = ((u \cos \tau - v \sin \tau) r + \dot{\theta}) r_\tau + (v \cos \tau + u \sin \tau) r^2 \tag{12}$$

5. Visual yaw detector

5.1. Moving observer

Knowing $f(t, \tau)$, $v = \text{const}$, and $u = 0$, we would like to estimate $\dot{\theta}$. We might expect the measured f to be rather noisy. Substituting the known values into equation 7 and considering some fixed particular value of τ gives a Riccati equation,

$$e(\dot{\theta}, \frac{d}{dt}\dot{\theta}, t; \tau) = A(t;\tau)\frac{d}{dt}\dot{\theta} + B(t;\tau)\dot{\theta}^2 + C(t;\tau)\dot{\theta} + D(t;\tau) = 0. \tag{13}$$

where

$$\begin{aligned} A(t;\tau) &= f_\tau^3 \sin\tau \\ B(t;\tau) &= -f_\tau^3 \cos\tau \\ C(t;\tau) &= 3f_\tau^2 f_t \cos\tau \\ D(t;\tau) &= -2f_t^2 f_\tau \cos\tau - (f_{tt}f_\tau^2 - 2f_t f_{t\tau} f_\tau + f_t^2 f_{\tau\tau})\sin\tau \end{aligned}$$

There are many such equations, corresponding to different values of τ. This is fortunate, since considered individually each one would be hopelessly corrupted by noise. To approximately solve them all simultaneously, we minimize the functional

$$\mathcal{E}(\dot{\theta}) = \int_{t_0}^{t_1} F(\dot{\theta}, \ddot{\theta}, t)\, dt \tag{14}$$

where[1]

$$F(\dot{\theta}, \ddot{\theta}, t) = I\left\{L(e(\dot{\theta}, \ddot{\theta}, t; \tau))\right\},$$

$L(\xi) = \xi^2$ is a loss function[2] and we define $I\{h(\tau)\} = \int_0^{2\pi} h(\tau)\, d\tau$. Note that $I\{\cdot\}$ is linear. Also note that $I\{(\partial/\partial\tau)h(\tau)\} = (\partial/\partial\tau)I\{h(\tau)\} = 0$, so for instance $I\{f_\tau\} = 0$ and $I\{ff_\tau\} = I\left\{(\partial/\partial\tau)f^2\right\}/2 = 0$. The extremal $\dot{\theta}(t)$ can be found using the calculus of variations. $\mathcal{E}$ is of the form required by Euler's equation, which states that the functional $\mathcal{E}$ is extremal when $\dot{\theta}$ satisfies

$$\frac{\partial F}{\partial \dot{\theta}} = \frac{d}{dt}\frac{\partial F}{\partial \ddot{\theta}}$$

Evaluating and simplifying this gives a second-order ordinary differential equation whose coefficients are easy to calculate numerically from visual data,

$$\begin{aligned} I\left\{A^2\right\}\frac{d^2}{dt^2}\dot{\theta} + I\left\{2\dot{A}A\right\}\frac{d}{dt}\dot{\theta} - I\left\{2B^2\right\}\dot{\theta}^3 + I\left\{\dot{A}B + A\dot{B} - 3BC\right\}\dot{\theta}^2 & \\ + I\left\{\dot{A}C + A\dot{C} - 2BD - C^2\right\}\dot{\theta} + I\left\{\dot{A}D + A\dot{D} - CD\right\} &= 0 \end{aligned} \tag{15}$$

[1] It would be reasonable to add a regularizer $\mathcal{R}(\dot{\theta})$ to F, to encourage $\dot{\theta}$ to be smooth.

[2] In the presence of occlusion and other local violations of equation 7 we would want to use a more robust estimator than simple squared error. Unfortunately this tends to make minimizing $\mathcal{E}$ more difficult. An ordinary differential equation still results if L is a polynomial, but higher degrees in L would make for higher powers of $\dot{\theta}$ and $\ddot{\theta}$ in equation 15.

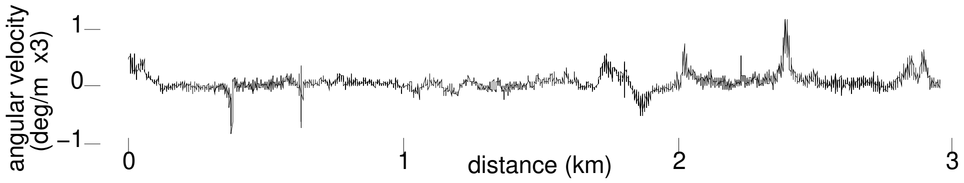


Figure 1 Equation 17 applied to a database of visual strips taken from around the horizon during a 3 km trek on a rural route, sampled every meter. Estimated $\dot{\theta}$ is shown as a function of distance. No smoothing has been done at any point in the processing, except that the strips were downsampled to 120 bins around the horizon.

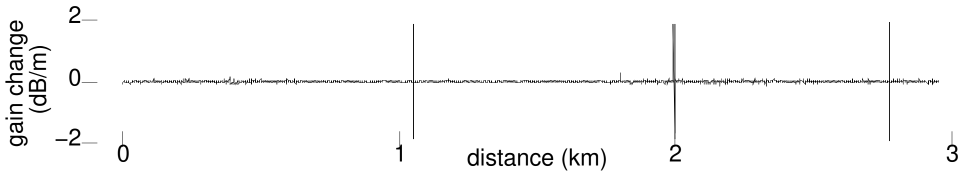


Figure 2 Equation 19 applied to a database of visual strips taken from around the horizon during a 3 km trek on a rural route, sampled every meter. Estimated log gain change is shown as a function of distance. No smoothing has been done at any point in the processing, except that the strips were downsampled to 120 bins around the horizon. Note the spikes, which correspond to glitches in the image acquisition process.

where $\dot{X} = (d/dt)X$.

An online ode solver, in which only the most recent few values of $\dot{\theta}$ are subject to change, can be used in settings that require running estimates of $\dot{\theta}$.

5.2. Stationary rotating observer

Equation 15 inherits an implicit assumption that $v \neq 0$, *i.e.* that the observer has non-negligible proper motion, from equation 7. Under some circumstances we might wish to make the opposite assumption. The analog of equation 7 when $u = v = 0$ is simply

$$f_t = \dot{\theta} f_\tau \tag{16}$$

Applying the calculus of variations as above with $F = I\left\{(\dot{\theta} f_\tau - f_t)^2\right\}$ gives[3] the trivial differential equation

$$\dot{\theta} = \frac{I\{f_t f_\tau\}}{I\{f_\tau^2\}} = \frac{\int_0^{2\pi} f_t f_\tau \, d\tau}{\int_0^{2\pi} f_\tau^2 \, d\tau} \tag{17}$$

An application of this equation is shown in figure 1.

---

[3]Instead of this F, we could instead choose a different F, say $F = I\left\{x(\dot{\theta} f_\tau - f_t)^2\right\}$ which leads to $\dot{\theta} = I\{x f_\tau f_t\} / I\left\{x f_\tau^2\right\}$. With $x = 1/f_\tau$ this gives $\dot{\theta} = I\{f_t\} / I\{f_\tau\}$. With $x = 1/f_t$ this gives $\dot{\theta} = I\{f_\tau\} / I\left\{f_\tau^2/f_t\right\}$. With $x = f_t^2 f_\tau^3 \cos\tau$ this gives $\dot{\theta} = I\left\{f_t^2 f_\tau^5 \cos\tau\right\} / I\left\{f_t^3 f_\tau^4 \cos\tau\right\}$. If there were no noise, these would all give the same $\dot{\theta}$. But with noise, the results will be different. How can we choose the best one? Assuming the noise in the measured f is independent zero mean Gaussian with constant variance, we should choose a quantity to integrate over both t and τ whose variance is the same everywhere. Within that constraint, we should choose the one with the lowest signal to noise ratio, in other words, the lowest ratio of standard deviation to mean. This is the F used in the text. A similar analysis must be applied to equation 14.

If the camera adds a constant extra signal to the entire visual field, as in a fluctuating background noise process, we can call this background noise $n_2(t)$ and model it as $f_t = \dot{\theta} f_\tau + n_2$. A more common type of noise is present in cameras with automatic gain control, which fluctuates due to factors outside the portion of the visual field being used here. Under these conditions, the change to a pixel's value attributable to camera gain change is proportional to that pixel's value. If the derivative of the log gain is $n_1(t)$ then process can be modeled as $f_t = \dot{\theta} f_\tau + n_1 f$. Combining these two noise processes yields

$$f_t = \dot{\theta} f_\tau + n_1 f + n_2 \tag{18}$$

Using $F = I\left\{(\dot{\theta} f_\tau - f_t + n_1 f + n_2)^2\right\}$ gives

$$\begin{aligned} \begin{pmatrix} \dot{\theta} \\ n_1 \\ n_2 \end{pmatrix} &= I\left\{ \begin{pmatrix} f_\tau \\ f \\ 1 \end{pmatrix} \begin{pmatrix} f_\tau \\ f \\ 1 \end{pmatrix}^T \right\}^{-1} I\left\{ f_t \begin{pmatrix} f_\tau \\ f \\ 1 \end{pmatrix} \right\} \\ &= \left(\begin{array}{c|cc} I\{f_\tau^2\} & & \\ \hline & I\{f^2\} & I\{f\} \\ & I\{f\} & I\{1\} \end{array} \right)^{-1} I\left\{ f_t \left(\begin{array}{c} f_\tau \\ \hline f \\ 1 \end{array} \right) \right\} \end{aligned} \tag{19}$$

which shows that the influences of these noise sources and the influence of rotation are orthogonal, and therefore the recovered $\dot{\theta}$ should not be sensitive to these two particular kinds of noise. However, the recovered gain changes $n_1(t)$ appear to detect camera frame acquisition errors quite well, as shown in figure 2.

6. Discussion

The above presentation is for an observer in flatland. This is the character of the application for which these techniques were developed, but the approach and equations extend to a full three dimensional setting. The only complication is the tedious algebra which arises from the coordinate systems that must be used on the surface of a sphere.

6.1. Scale

In many of the equations above, the velocity appears as $\dot{v}/v = (d/dt)\log v$. It is clear from the physics of the situation that this must be so, for even if everything but distances are known, the velocity can be determined uniquely only up to a scale factor. In other words, if $v(t)$ is a solution to the equation, then so must be $\alpha v(t)$ for any $\alpha > 0$, because $(d/dt)\log(\alpha v) = (d/dt)\log v$. This is of course common in computer vision.

6.2. Other approaches

There are many approaches to recovering egomotion or distances from image sequences.

We have taken a gradient-based approach, similar to that of Horn and Weldon (1988), Negahdaripour and Horn (1987), in contrast to the alternative correlation based approaches (Hassenstein and Reichardt, 1956; Poggio and Reichardt, 1973). Gradients are more susceptible to noise than correlations, and present difficulties when the spatial and temporal discretization imposed by modern digital computer technology makes for image movements greater than a few pixels per time step (Hildreth and Koch, 1987). This latter problem can be ameliorated by spatially downsampling the image, or, since often the ideal sampling rate varies across a single image, by adaptive multiscale techniques. On the other hand, there are theoretical reasons to believe that gradient methods perform better than correlation methods in the high signal to noise regime (Potters and Bialek, 1994).

Another popular indirect method is point tracking, as in the matrix decomposition method of Tomasi and Kanade (1992) of the chronogeneous motion algorithm of Franzen (1991).

The approach taken here is a direct one, since image gradients are used directly, rather than being used to calculate the optical flow (Horn and Schunck, 1981; Poggio, Yang, and Torre, 1989; Verri and Poggio, 1989) which in turn constitutes the input to an egomotion module (Uras *et al.*, 1989).

6.3. Noise model

The equations shown in the text are optimized for a small amount of additive gaussian noise, and fall into the MLE framework, since they are derived from a least squared formulation and incorporate no regularizer or prior on the egomotion. Unfortunately that is not the sort of noise actually encountered in practice. Instead, there are a few sources of noise, of which, at least for our applications, Gaussian camera noise is the least significant. More severe are lack of stationarity of the world, such as other moving vehicles; occlusion and revelation; specularities and reflections; and camera blooming from the sun and reflections thereof.

Also, an additive gaussian noise model is a poor model of the world. In essence, it assumes that an object's intensity and distance follow a random walk. Instead, it is our observations which are corruptions of an underlying unchanging stationary object.

Some of these problems could be dealt with in part by complicating the model, for instance including occlusion processes which flow around the visual field according to the same equation by which objects do. These could be created by the broken spring models so popular computer vision, which of course correspond to robust estimators, which themselves might improve performance (Poggio, Torre, and Koch, 1985; Koch, 1988).

Another route for improvement would be to attempt to maintain a world model *via* an occupancy grid (Moravec, 1988). This would have the added benefit that it could be matched against a database, hopefully making the system more robust to seasonal variation in the visual appearance of objects. Unfortunately the computational burden might exceed the capacities of our target platform.

6.4. Color

Lastly, the current system does not make use of color information. Since color can be more stable to view angle than intensity, it would make sense to incorporate color in a non-trivial fashion into a system capable of detecting long range motion. However here we use a gradient method, which is will operate properly only for small changes to the visual field, on the order of a pixel or less. It is interesting to note that in primates short range motion makes little use of color, while long range motion is quite sensitive to color boundaries (Ramachandran and Gregory, 1978).

References

Franzen, W. O. (1991). Structure and motion from uniform 3D acceleration. In *Proceedings of the IEEE Workshop on Visual Motion*, pp. 14–20 Princeton, NJ. IEEE Press.

Gorr, R. E., Hancock, T. R., Judd, J. S., Lin, L.-J., Novak, C. L., and Rickard, Jr., S. T. (1995). A Vision Based System for Automatic Vehicle Location. Tech. rep. SCR-95-TR-534, Siemens Corporate Research, Princeton, NJ.

Hassenstein, B. and Reichardt, W. (1956). Systemtheoretische Analyse der Zeit, Reihenfolgen, und Vorzeichenauswertung bei der Bewegungsperzepion des Rüsselkäfers *Chlorophanus*. *Z. Naturforsch*, *11b*, 513–524.

Hildreth, E. C. and Koch, C. (1987). The Analysis of Visual Motion: from Computational Theory to Neuronal Mechanisms. *Annual Review of Neuroscience*, *10*, 477–533.

Horn, B. K. P. and Schunck, B. G. (1981). Determining Optical Flow. *Artificial Intelligence*, *17*, 185–203.

Horn, B. K. P. and Weldon, Jr., E. J. (1988). Direct Methods for Recovering Motion. *International Journal of Computer Vision*, *2*, 51–76.

Koch, C. (1988). Computing Motion in the Presence of Discontinuities: Algorithm and Analog Networks. In Eckmiller, R. and von der Malsburg, C. (Eds.), *Neural Computers*. Springer-Verlag.

Lin, L.-J. and Judd, J. S. (1995). A Robust Landmark-Based System For Vehicle Location Using Low-Bandwidth Vision. Tech. rep. SCR-95-TR-535, Siemens Corporate Research, Princeton, NJ.

Moravec, H. P. (1988). Sensor Fusion in Certainty Grids for Mobile Robots. *AI magazine*, *9*(2), 61–74.

Negahdaripour, S. and Horn, B. K. P. (1987). Direct Passive Navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*(1), 168–176.

Novak, C. L. and Hancock, T. R. (1995). Probability Modelling for Visual Matching in Automatic Vehicle Location. Tech. rep. SCR-95-TR-532, Siemens Corporate Research, Princeton, NJ.

Novak, C. L., Hancock, T. R., Rickard, Jr., S. T., and Judd, J. S. (1995). A Visual Gyroscope for Automatic Vehicle Location. Tech. rep. SCR-95-TR-531, Siemens Corporate Research, Princeton, NJ.

Poggio, T. and Reichardt, W. (1973). Considerations on models of movement detection. *Biological Cybernetics*, *13*, 223–227.

Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*(6035), 314–319.

Poggio, T., Yang, W., and Torre, V. (1989). *Optical Flow: Computational Properties and Networks, Biological and Analog*. Addison-Wesley.

Potters, M. and Bialek, W. (1994). Statistical mechanics and visual signal processing. *J. Phys. I France*, *4*, 1755–1775.

Ramachandran, V. S. and Gregory, R. L. (1978). Does color provide an input the human motion perception?. *Nature*, *275*, 55–56.

Tomasi, C. and Kanade, T. (1992). Shape and Motion from Image Streams under Orthography. *International Journal of Computer Vision*, *9*(2), 137–154.

Uras, S., Girosi, F., Verri, A., and Torre, V. (1989). A computational approach to motion perception. *Biological Cybernetics*, *60*, 79–87.

Verri, A. and Poggio, T. (1989). Motion Field and Optical Flow: Qualitative Properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 490–498.