[6]

Estimation of Stereo and Motion Parameters using a Variational Principle

Harit P Trivedi

BP Research Centre
Sunbury-on-Thames, TW16 7LN, UK

Reprinted, with permission of Butterworth Scientific Ltd, from *Image and Vision Computing*, 1987, **5**, 181-183.

The problems of extracting 3D structure from stereo or motion parameters from optic flow are now analytically tractable but numerically ill-conditioned. A variational principle is proposed which alleviates ill-conditioning and saturates rapidly with data so that even a small excess (over a minimal number) of data points yields accurate results. It involves no adjustable parameters (unlike many applications of the regularization theory) and no assumptions about measurement errors, which, in fact, it seeks to estimate and minimize. The technique is illustrated with image resolutions varying from 1024 to 128 pixels square, using between 6 and 30 data points (5 data points define a unique solution) perturbed by at most 0.2 pixels. The error in the computed direction of translation was 2.7 deg in the worst case (128 x 128 pixels, 15 data points). It was 1.2 deg with only six data points for an image 1024 square.

Keywords: stereo vision, motion, optic flow, variational principle

The low-level computer vision problems of determining camera geometry from stereo images and object motion and structure from optic flow (or a time sequence of images) are ill-conditioned. Tsai and Hyang (1984), for example, reported a staggering 54% error in the model parameters for only a 1% error in the data. Fang and Huang (1984) also observed similar symptoms of numerical instability.

Researchers have responded to this difficulty (for example, Yasumoto and Medioni, 1985) by using the regularization techique (Tikhonov and Arsenin, 1977), following the lead of Poggio (Poggio T, 1985). This involves additional constraints, which are quite often heuristic, and each constraint entails a weighting coefficient, which, in practice at least, has to be chosen judiciously if not with some degree of foreknowledge.

In this paper, a variational principle is formulated which involves no additional constraints, has no adjustable parameters (such as weighting coefficients), 'corrects' each data point, and yields errors entirely in terms of the image measurables. As one usually knows something about the accuracy of the device, the precision of edge-location, etc, one can judge the quality of the computed solution from the estimated measurement error.

VARIATIONAL PRINCIPLE

Let the model equation be $f(\mathbf{x}, \mathbf{a}) = 0$, where $\mathbf{x}$ denotes a data point and $\mathbf{a}$ denotes the model parameters. Allowing for error in the data measurements, a correction $\delta\mathbf{x}(i)$ is sought to satisfy exactly

$$f[\mathbf{x}(i) + \delta\mathbf{x}(i), \mathbf{a}] = 0 \qquad \text{for } i = 1, 2, ..., N \qquad (1)$$

This will not determine $\delta\mathbf{x}(i)$ uniquely, of course, and so we impose a subsidiary condition and select that $\delta\mathbf{x}(i)$ in each instance i for which the size of the correction

$$|\delta\mathbf{x}(i)|^2 = [\delta x_1(i)]^2 + ... + [\delta x_m(i)]^2 \qquad (2)$$

is the smallest. That is,

$$f[\mathbf{x}(i) + \delta\mathbf{x}(i),\mathbf{a}] = 0 \qquad |\delta\mathbf{x}(i)|^2 \text{ minimum } \forall\, i \qquad (3)$$

By defining the size of the correction by Equation (2), all the components of a data measurement have been put on an equal footing. It is also implied that the absolute correction is the most meaningful. While this is so in our application domain (the measured data are the coordinates of image points), these assumptions are not necessarily universal, and, in general, the size of the correction must be defined in a way appropriate to the problem at hand. If the $\delta\mathbf{x}$ values are required to obey constraints, the equation above must be solved subject to those constraints.

Let the kth component of the formal solution of Equation (3) be written as

$$\delta x_k(i) = g_k[\mathbf{x}(i), \mathbf{a}] \qquad k === 1, ..., m \qquad (4)$$

Neglecting second- and higher-order terms in the Taylor series of expansion of f in Equation (3) above, it can be determined that

$$\mathbf{g}(\mathbf{x},\mathbf{a}) = - f\nabla f/|\nabla f|^2 \qquad (5)$$

Then

$$E(i) = |\delta x(i)|^2 = \sum_{k=1}^{m} g_k[x(i), a]^2 \qquad (6)$$

is the minimum correction to $\mathbf{x}(i)$ given $\mathbf{a}$, and

$$e = \sum_{i=1}^{N} E(i) \qquad (7)$$

is the minimum total correction to the sampled data, given $\mathbf{a}$. This, then, is the varational quantity to be minimized with respect to the latter, the model parameters. It generates that solution which, for the smallest correction to the sampled data, enables the model equation to be satisfied exactly at each (corrected) data point.

Application of variational principle to sample problem

No of data points	Image resolution	1 - ê.ê'	$\left[\sum_i (\hat{e}_i - \hat{e}_i')^2\right]^{\frac{1}{2}}$	$\cos^{-1}$(ê.ê') (deg)	1 - **T.T'**	$\left[\sum_i (T_i - T_i')^2\right]^{\frac{1}{2}}$	$\cos^{-1}$(**T.T'**) (deg)
6	1024^2	2.4×10^{-7}	6.9×10^{-4}	0.040	2.3×10^{-4}	5.0×10^{-3}	1.2
7	1024^2	2.7×10^{-7}	7.3×10^{-4}	0.042	2.3×10^{-4}	5.0×10^{-3}	1.2
8	1024^2	3.3×10^{-7}	8.1×10^{-4}	0.047	1.1×10^{-4}	3.4×10^{-3}	0.85
9	1024^2	2.5×10^{-7}	7.1×10^{-4}	0.040	4.9×10^{-5}	2.2×10^{-3}	0.57
10	1024^2	1.7×10^{-7}	5.9×10^{-4}	0.034	1.6×10^{-6}	1.0×10^{-5}	0.10
15	1024^2	1.3×10^{-7}	5.0×10^{-4}	0.029	2.1×10^{-5}	1.4×10^{-3}	0.37
30	1024^2	1.7×10^{-9}	5.9×10^{-5}	0.0034	6.5×10^{-7}	2.4×10^{-4}	0.066
15	512^2	6.3×10^{-7}	1.1×10^{-3}	0.064	1.9×10^{-5}	1.4×10^{-3}	0.36
15	256^2	3.4×10^{-6}	2.6×10^{-3}	0.15	1.2×10^{-4}	3.0×10^{-3}	0.9
15	128^2	9.2×10^{-6}	4.3×10^{-3}	0.25	1.1×10^{-3}	9.3×10^{-3}	2.7

Computed solution (ê',T') and true solution (ê,T): [ê.ê = ê'.ê' = 1, **T.T** = **TT'** = 1]
An image is a unit square, unit distance from the optic centre. The error is a uniformly distributed random value bounded by $|\Delta x| < 0.2$ pixels, $|\Delta y| < 0.2$ pixels. As a result, the absolute of the error doubles as the image resolution halves (from 1024 value to 512 etc). The results correspond to $T_1{:}T_2{:}T_3 = 1{:}0.1{:}0.2$ and the rotation $R = R_z(0.03)R_y(0.2)R_z(0.05)$, the angles being in radians. The depth varied between 5 and 100 interocular units.

EXAMPLE

Image coordinates **x** and **x'** in the left and right images of a stereo pair corresponding to every scene point obey the relation[1,6,7]

$$\sum_{i,j=1}^{3} x_i' Q_{ij} x_j = 0 \tag{8}$$

where x'_3, $x_3 = 1$. The primed coordinates refer to the right image and the unprimed to the left. The matrix **Q** = **RS** is defined in terms of the rotation matrix **R** and the antisymmetric matrix **S** related to the translation vector **T** = (T_1, T_2, T_3) by

$$S = \begin{bmatrix} 0 & T_3 & -T_2 \\ -T_3 & 0 & T_1 \\ T_2 & -T_1 & 0 \end{bmatrix} \tag{9}$$

perspective projection being assumed.

The aim is to determine the matrices **R** and **S** of the stereo geometry, given N data points. This is found to be an ill-conditioned problem. That is, errors in the data are amplified when the model parameters are determined from the (imperfect) data. We have used g of Equation (5) to apply our method to this problem. Each point was perturbed by a random amount (both horizontally and vertically) bounded by ± 0.2 pixel. The distribution of pertubation over the points was uniform. The results are summarized in Table 1. The rotation matrix was parametrized using Euler parameters[8], which are real unlike the Cayley-Klein parameters) and are distinct from Euler angles. For completeness,

$$R = \begin{bmatrix} e_0^2 + e_1^2 - e_2^2 - e_3^2 & 2(e_1 e_2 + e_0 e_3) & 2(e_1 e_3 - e_0 e_2) \\ 2(e_1 e_2 - e_0 e_3) & e_0^2 - e_1^2 + e_2^2 - e_3^2 & 2(e_2 e_3 + e_0 e_1) \\ 2(e_1 e_3 + e_0 e_2) & 2(e_2 e_3 - e_0 e_1) & e_0^2 - e_1^2 - e_2^2 + e_3^2 \end{bmatrix} \tag{10}$$

where $e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1$. Since Equation 8 clearly leaves the overall scale of **T** undetermined, it was fixed by setting **T.T** = 1.

RESULTS AND CONCLUSIONS

A variational principle, the minimum correction principle, has been constructed to deal with ill-conditioned problems. A minimum correction to the sampled data is sought, such that the corrected data obeys exactly the model equation in each instance (or, to be precise, through the first order in the corrections, at least).

Unlike some applications of the regularization theory, the minimum correction principle involves no adjustable parameters. In fact, it seeks to estimate data errors by minimizing them with respect to the model parameters and requires no assumptions to be made about them. The principle has been illustrated with various image resolutions, from 1024 to 128 pixels square, using between six and 30 data points perturbed by at most 0.2 pixels. The error in the computed direction of translation was 2.7 deg in the worst case (image resolution 128 x 128 with 15 data points. It was 1.2 deg with only six data points for an image 1024 square. (It should be recalled that five data points are needed to even define a solution.) The accuracy of the rotational parameters is much greater, as observed by Tsai and Huang (1984). (There is an explanation for this behaviour, although it is not directly

Table 2. Comparison of least squares method (singular value decomposition based) and minimum correction principle for different field of view angles.

Resolution	No of matches	tan θ = 0.5 Least sq	tan θ = 0.5 Min corr	tan θ = 0.25 Least sq	tan θ = 0.25 Min corr
1024	6	-	1.2	-	1.8
1024	7	-	1.2	-	1.7
1024	8	-	0.9	-	1.2
1024	9	64	0.6	69	1.0
1024	10	1.7	0.1	7.7	1.1
1024	15	0.3	0.4	2.7	1.2
1024	30	0.5	0.07	0.4	0.05
512	15	0.3	0.4	1.0	1.1
256	15	10.0	0.9	74	1.1
128	15	64	2.7	141	1.0

[tan θ = (1/2) image width (or height)/focal length]. Camera geometry and data errors are as in Table 1.

connected with this work. The variation with respect to **T**, subject to a normalization constraint, can be performed analytically in both methods. The result is that **T** has to be an eigenvector of a matrix depending on the data measurements and **R**; hence the observed behaviour.) The four Euler parameters are characterized by a direction in a 4D Euclidean space. The error in the computed direction (of rotation) was found to be minuscule (0.25 deg at most). No compelling theoretical grounds have yet been found that might explain why this method works so well. To this extent, it remains an empirical method.

The method is numerically stable and saturates rapidly, so that a single extra data point usually suffices. It is insensitive to the addition or removal of data points (stability). The numerical value of the functional E is the estimated total error in the sampled data. For 'true' model parameters, the correction to each component of data is just the negative of the error in its measured value. Knowing the precision of the data measurements *a priori* (from factors like resolution, quantization, device accuracy, etc), one can judge the solution quality by comparing these two quantities. For example, a higher than expenced value of $E(i)$ can expose a rogue data point.

When the exact solution of the model Equation (5) with corrected data is impracticable, a linear approximation (keeping only constants and linear terms in $\delta \mathbf{x}$) can be made. The stereo and motion results above were obtained using this approximation.

Our error measures are independent of the laboratory coordinate frame. To compare the performance of the varational principle with other methods, a conventional least-squares calculation (using singular value decomposition) was performed for two different field-of-view angles. The results are compared in Table 2. While the variational method produces smaller errors, the more striking finding is that the results of the least-squares method look far bettwer than one might have expected, given the notoriety of this particular problem. The importance of choosing appropriate (ie coordinate frame invariant) error measures to evaluate and compare methods cannot be overemphasized.

Although this method entails more work, it seems to handle adequately the ill-conditioned problem of determining stereo geometry even with a single extra data point. Certain 'degenerate' configurations of data points, as pointed out by Longuet-Higgins (1984) and Tsai and Huang (1984), cause the '8-point algorithm' [1,7] for solving for the eight ratios of Q in Equation (8) to break down. The method described here does not suffer from this problem as the variation is performed directly with respect to the parameters of rotation and translation.

ACKNOWLEDGEMENTS

It is a pleasure to thank Bernard Buxton for the many useful discussions during the course of this work.

REFERENCES

Tsai, R. Y. and Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. Pattern Anal. & Mach. Intell.* **6**, 13-26.

Fang, J-Q. and Huang, T.S. (1984). Some experiments on estimating 3D motion parameters of a rigid body from two consecutive image frames. *IEEE Trans. Pattern Anal. & Mach. Intell.* **6**, 545-554.

Yasumoto, Y. and Medioni, G. (1985). Experiments in estimation of 3D motion parameters from a sequence of image frames. *Proc. I.E.E.E. Conference on Computer Vision & Pattern Recognition*, San Francisco, CA, U.S.A., 89-94.

Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of ill-posed problems.* V.H. Winston, Washington, DC. U.S.A.

Poggio, T. (1985) Early vision: from computational structure to algorithms and parallel hardware. *Computer Vision, Graphics Image Proc.*, **31**, 139-155.

Thompson, E.H. (1959) A rational algebraic formulation of the problem of relative orientation *Photogramm. Record*, **3**, 152-159 (especialy Note 2).

Longuet-Higgins, H.C. (1981) A computer algorithm for reconstructing a scene from two projections. *Nature* **293**, 133-135.

Goldstein, H. (1980) *Classical mechanics (2nd edn)* Addison-Wesley, Reading, MA, U.S.A., p.153 and appendix B.

Longuet-Higgins, H.C. (1984) The reconstruction of a scene from two projections - configurations that defeat the 8-point algorithm. *Proc. 1st Conf. Appl. of AI*, Denver, CO, U.S.A. 395-397.