Estimation of Stereo and Motion Parameters using a Variational Principle

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


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

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

$$\| \delta x(i) \|^2 = [\delta x_1(i)]^2 + \ldots + [\delta x_m(i)]^2 \quad (2)$$

is the smallest. That is,

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

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 $\delta x$ values are required to obey constraints, the equation above must be solved subject to those constraints.

Let the $k$th component of the formal solution of Equation (3) be written as

$$\delta x_k(i) = g_k[x(i), a] \quad k = 1, \ldots, m \quad (4)$$

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

$$g(x, a) = - f / \sqrt{f / f^2} \quad (5)$$

Then

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

is the minimum correction to $x(i)$ given $a$, and

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

is the minimum total correction to the sampled data, given $a$. 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

<table>
<thead>
<tr>
<th>No of data points</th>
<th>Image resolution</th>
<th>1 - ( e \cdot e' )</th>
<th>( \sum (e_i - e'_i)^2 )</th>
<th>( \cos^{-1}(e \cdot e') )</th>
<th>1 - T( T' )</th>
<th>( \sum (T_i - T'_i)^2 )</th>
<th>( \cos^{-1}(T \cdot T') )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>2.4 \times 10^{-7}</td>
<td>6.9 \times 10^{-4}</td>
<td>0.040</td>
<td>2.3 \times 10^{-4}</td>
<td>5.0 \times 10^{-5}</td>
<td>1.2</td>
</tr>
<tr>
<td>7</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>2.7 \times 10^{-7}</td>
<td>7.3 \times 10^{-4}</td>
<td>0.042</td>
<td>2.3 \times 10^{-4}</td>
<td>5.0 \times 10^{-5}</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>3.3 \times 10^{-7}</td>
<td>8.1 \times 10^{-4}</td>
<td>0.047</td>
<td>1.1 \times 10^{-4}</td>
<td>3.4 \times 10^{-5}</td>
<td>0.85</td>
</tr>
<tr>
<td>9</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>2.5 \times 10^{-7}</td>
<td>7.1 \times 10^{-4}</td>
<td>0.040</td>
<td>4.9 \times 10^{-5}</td>
<td>2.2 \times 10^{-5}</td>
<td>0.57</td>
</tr>
<tr>
<td>10</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1.7 \times 10^{-7}</td>
<td>5.9 \times 10^{-4}</td>
<td>0.034</td>
<td>1.6 \times 10^{-6}</td>
<td>1.0 \times 10^{-5}</td>
<td>0.10</td>
</tr>
<tr>
<td>15</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1.3 \times 10^{-7}</td>
<td>5.0 \times 10^{-4}</td>
<td>0.029</td>
<td>2.1 \times 10^{-5}</td>
<td>1.4 \times 10^{-3}</td>
<td>0.37</td>
</tr>
<tr>
<td>30</td>
<td>1024&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1.7 \times 10^{-9}</td>
<td>6.5 \times 10^{-7}</td>
<td>0.0034</td>
<td>6.5 \times 10^{-7}</td>
<td>2.4 \times 10^{-4}</td>
<td>0.066</td>
</tr>
<tr>
<td>15</td>
<td>512&lt;sup&gt;2&lt;/sup&gt;</td>
<td>6.3 \times 10^{-7}</td>
<td>1.1 \times 10^{-3}</td>
<td>0.064</td>
<td>1.9 \times 10^{-5}</td>
<td>1.4 \times 10^{-3}</td>
<td>0.36</td>
</tr>
<tr>
<td>15</td>
<td>256&lt;sup&gt;2&lt;/sup&gt;</td>
<td>3.4 \times 10^{-6}</td>
<td>2.6 \times 10^{-3}</td>
<td>0.15</td>
<td>1.2 \times 10^{-4}</td>
<td>3.0 \times 10^{-3}</td>
<td>0.9</td>
</tr>
<tr>
<td>15</td>
<td>128&lt;sup&gt;2&lt;/sup&gt;</td>
<td>9.2 \times 10^{-6}</td>
<td>4.3 \times 10^{-3}</td>
<td>0.25</td>
<td>1.1 \times 10^{-3}</td>
<td>9.3 \times 10^{-3}</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Computed solution \((e', T')\) and true solution \((e, T)\): \[ e.e' = e'.e' = 1, \quad T.T = TT' = 1 \]

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

\[
\sum_{i,j=1}^{3} x'_i Q_{ij} x_j = 0
\]

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

\[
S = \begin{bmatrix}
0 & T_3 & -T_2 \\
-T_3 & 0 & T_1 \\
T_2 & -T_1 & 0
\end{bmatrix}
\]

\[ R = \begin{bmatrix}
e_1^2 + e_2^2 - e_3^2 & 2(e_1 e_2 + e_0 e_3) & 2(e_1 e_3 - e_0 e_2) \\
2(e_1 e_2 - e_0 e_3) & e_3^2 - e_1^2 + e_2^2 & 2(e_2 e_3 + e_0 e_1) \\
2(e_1 e_3 + e_0 e_2) & 2(e_2 e_3 - e_0 e_1) & e_0^2 - e_1^2 - e_2^2 + e_3^2
\end{bmatrix}
\]

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.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>No of matches</th>
<th>tan θ = 0.5</th>
<th>tan θ = 0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Least sq</td>
<td>Min corr</td>
<td>Least sq</td>
</tr>
<tr>
<td>1024</td>
<td>6</td>
<td>-</td>
<td>1.2</td>
</tr>
<tr>
<td>1024</td>
<td>7</td>
<td>-</td>
<td>1.2</td>
</tr>
<tr>
<td>1024</td>
<td>8</td>
<td>-</td>
<td>0.9</td>
</tr>
<tr>
<td>1024</td>
<td>9</td>
<td>64</td>
<td>0.6</td>
</tr>
<tr>
<td>1024</td>
<td>10</td>
<td>1.7</td>
<td>0.1</td>
</tr>
<tr>
<td>1024</td>
<td>15</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>1024</td>
<td>30</td>
<td>0.5</td>
<td>0.07</td>
</tr>
<tr>
<td>512</td>
<td>15</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>256</td>
<td>15</td>
<td>10.0</td>
<td>0.9</td>
</tr>
<tr>
<td>128</td>
<td>15</td>
<td>64</td>
<td>2.7</td>
</tr>
</tbody>
</table>

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

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


