Robust Estimation of 3-D Motion Parameters in Presence of Correspondence Mismatches

Subhasis Chaudhuri
Dept. of Electrical Engg.
Indian Institute of Technology
Powai, Maharashtra 400 076
INDIA

Abstract

Accuracy of a motion analysis scheme based on feature point correspondence is quite poor when there are mismatches between points. A small amount of mismatch (i.e., outlier data) may degrade the performance of the least-squares estimator significantly. Least median of squares (LMedS) estimator can provide the required robustness. This method works well even when almost half of the given data are outliers. A Monte-Carlo sampling technique is used to reduce the computational load. Subsequent use of the total least-squares (TLS) or the constrained least-squares (CLS) method on the trimmed data set enhances efficiency. Simulation experiments are carried out to evaluate the performance of the proposed method.

1 Introduction

The motion of a rigid object may be given by an affine transform

$$\bar{z}_i = \bar{x}_i R + \bar{t}$$

where $\bar{x}_i$ is a row vector representing the coordinates of the ith point in the object, $\bar{z}_i$ is the coordinates of the same point after the motion, $R$ is a 3 x 3 orthonormal rotation matrix, $\bar{t} = [tx ty tz]$ is the translation of the object between two successive time frames. The motion understanding problem involves the estimation of the orthonormal matrix $R$ and the vector $t$ in presence of sensor noise, [1]. Performances of the motion estimation methods depend on the accuracy with which the point correspondences have been established. Accuracy of these algorithms when there are mismatches in point correspondence has been investigated in [2], [3].

In range imaging, 3-D coordinates of feature points on the object are assumed to be available in an object centered coordinate system. The translation vector $t$ can be easily estimated by noting the change in the centroid of the object in two successive frames, provided there is no occlusion. Thus, one needs to estimate only the rotation matrix $R$, when the object is assumed to have undergone a rotation $\psi$ about an axis, passing through the centroid of the object, with direction cosines $(n_1, n_2, n_3)$. Given point correspondences, a constrained least squares estimate of $R$ can be obtained [4], which imposes the orthonormality of the estimated $R$ matrix. This implicitly assume the measurement errors confined to a single frame only. However, both frames are prone to measurement noise. To account for the noise in both time frames, the use of the total least squares (TLS) method has been suggested [5]. Although the TLS solution does not guarantee orthonormality, it is quite suitable to deal with an object undergoing simultaneous linear deformation (pseudo-rigid). Furthermore, the CLS method is not applicable in estimating motion parameters from visual images, but the TLS method can still be used.

Let $T$ be a regression estimator on the data set $Z = \{z_i\}_{i=1}^N$, and $Z'$ denote all possible corrupted samples obtained by replacing any $m$ (outliers) of the original data. The maximum bias caused by such outliers is given by

$$\text{bias}_N(m; T, Z) = \sup_{Z'} ||T(Z') - T(Z)||.$$}

The breakdown point of the estimator $T$ is defined as [6],

$$\epsilon^*_N(T) = \min_{m} \left\{ \frac{m}{N}; \text{bias}_N(m; T, Z) \text{ is infinite} \right\},$$

and is independent of the probability distribution of the outlier. It may be noted that for the least squares method, a single bad outlier is sufficient to ‘break down’ the estimator (i.e., $\epsilon^* = 1/N$). Numerical studies about the degradation of the performance of the LS estimators in presence of feature point mismatch have been reported in [2], [3].
Recently, Rousseeuw has suggested a new robust paradigm for linear regression, [6], [7], called the least median of squares (LMedS) estimator. It has been used in cluster analysis [8], in obtaining robust estimate of an autoregressive model [9], etc. The LMedS method has an asymptotic breakdown point $c^* = 0.5$, which is the maximum possible value that an estimator can achieve. Another advantage of this method is that one does not have to construct any influence function.

The organization of this paper is as follows. Least squares based methods are briefly presented in section 2. These LS methods will be used in conjunction with the LMedS estimator in section 3 to derive the motion analysis scheme. A brief sketch of the LMedS estimator, and the computational complexity of this method are presented in section 3. A Monte-Carlo sampling technique is used to speed up the computation. The results of the simulation experiments for range data as well as visual images are given in section 4. The merits and limitations of the proposed motion analysis scheme are discussed in section 5.

2 LS Estimation of Motion Parameters

2.1 Range Data

Let $N$ be the number of feature points being considered. Assuming an object centered coordinate system, the motion equation can be given by $Q = PR + \delta Q$, where $Q$ and $P$ are $N \times 3$ matrices representing the coordinates of the feature points in two successive frames, and $\delta Q$ is the corresponding perturbations in the $Q$ matrix. An LS estimate of the rotation matrix is given by $\hat{R} = (P^T P)^{-1} P^T Q$. For a rigid object, the CLS estimate of the rotation matrix is given by

$$\hat{R}_{CLS} = V U^T, \quad P^T Q = U \Sigma V^T. \quad (1)$$

When the data in both frames are corrupted by sensor noise, we may use the total least squares (TLS) method proposed in [10], [11], [12]. The closed-form solution can be obtained as follows [13], [5] : obtain a singular value decomposition (SVD) of the augmented matrix $[P \mid Q]$

$$[P \mid Q] = U \Sigma V^T. \quad (2)$$

Here $U$ and $V$ are of dimensions $N \times 6$ and $6 \times 6$ respectively. For the noisy observation, a rank-3 approximation is considered. We may write $V$ in the partitioned form

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad \text{each } V_{i,j} \text{ is } 3 \times 3$$

The TLS solution is given by

$$\hat{R}_{TLS} = -V_{12} V_{22}^{-1}. \quad (3)$$

Constrained TLS is identical to the CLS under the assumption that the error covariance matrices in both frames have identical structure.

2.2 Perspective Projection

Let $(x, y, z)$ and $(x', y', z')$ represent the coordinates of the 3-D feature points, and $(X, Y)$ and $(X', Y')$ represent the corresponding points on the image plane in two successive time frames respectively. From geometry, we may obtain, [14]

$$[X' \ Y'] = [X \ Y 1]^T = 0 \quad (4)$$

where $E = GR$, and

$$G = \begin{bmatrix} 0 & -t_x & t_y \\ t_x & 0 & -t_z \\ -t_y & t_z & 0 \end{bmatrix}.$$  

Eq.(4) is linear and homogeneous in the nine unknowns in $E$ matrix. Thus, the translation parameters can be estimated only up to a scale factor. Denoting $\alpha_j = \sqrt{(X_j^2 + Y_j^2 + 1)(X_j'^2 + Y_j'^2 + 1)}$, eq.(4) for each feature point, $j$, can be explicitly represented as

$$\frac{1}{\alpha_j} [X_j X_j' \ X_j Y_j' \ X_j Y_j \ Y_j' \ Y_j \ Y_j'] = [e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9]^T = 0, \quad (5)$$

where $e_i$'s are the elements of the $E$ matrix. This can be written as $A\hat{e} = 0$, where $A$ is an $N \times 9$ matrix. The LS solution involves perturbing the last column of the equation only so that it lies within the span of the data in first eight columns. The TLS solution acknowledges the possible presence of perturbations in all nine columns of the matrix $A$. If $A = U \Sigma V^T$, then $\hat{e}_{TLS}$ is given by the right singular vector corresponding to the smallest singular value.

3 Least Median of Squares Estimate

Let $z_i = \{p_i, q_i\}$ be the measurements of the $i$th feature point on the object. If $R$ is an estimate of the rotation matrix, the residual for the $i$th feature point is defined as $r_i = |q_i - p_i R|$ (or $r_i = |X_i Y_i'| E X_i Y_i'|^T / \alpha_i$ for visual image data).

In LMedS methods, the parameter $\theta$ (R or E matrix) is estimated by solving

$$\phi(\theta) = \min_{i} \mathop{\text{med}}_{i} r_i^2 \quad i = 1, 2, \ldots, N. \quad (6)$$

This estimator yields the smallest value of the median of the squared residuals for the entire data set. However, there is no analytic solution to eq. (6).
To solve the above, distinct \( k \)-tuples of data are chosen from the given data set, \( \{ z_1, \ldots, z_m \} \), where \( k \geq \text{dim}(\theta) \). There are \( \binom{N}{k} \) different choices. A trial estimate of \( \theta \) is computed for each. If \( k = \text{dim}(\theta) \), then an exact solution for the parameter vector \( \theta \) can be computed. Otherwise, the trial estimate \( \theta \) may be computed in a least squares sense. Here we use CLS for the range data and TLS for the image data. The median of the squared residuals \( r_j(\theta) \) for all the feature points is obtained. Then, for all possible \( k \)-tuples, one finds the smallest value of the median of residuals. The trial estimate that corresponds to the LMedS residual is the true LMedS estimate.

### 3.1 Nonuniqueness of LMedS

It is interesting to note that, the LMedS estimator may not yield a unique solution, even in absence of any outlier or sensor noise. If at least half of the data points \( m \geq \lceil N/2 \rceil + 1 \) lie in a hyperplane having dimension less than that of the parameter vector \( \theta \), then there is no unique estimate for \( \theta \). This is due to the fact that for each \( k \)-tuple chosen from such a subset of observations, there is no unique solution to the trial estimate as the observation matrix is rank deficient. This is analogous to the notion of sufficiently exciting (SE) signal in control theory. However, such a singular case in computer vision is unlikely.

### 3.2 Computational Issues

Computational complexity of the LMedS estimator is of the order of \( O(N^{k+1}) \). And the memory requirement is of the order of \( (\text{dim}(\theta) + 1)O(N^k) \). Space and time efficient algorithms for the LMedS estimator have been suggested in [15]. These algorithms deal with fitting a straight line through the given set of data.

The computation of the LMedS estimator for higher dimension, may be prohibitive. It increases exponentially with the increase in the size of \( k \)-tuple. In practice, one resorts to a Monte-Carlo sampling.

### 3.3 Sampling by Monte-Carlo Method

Here the probability of error \( P_e \) refers to the likelihood that one does not find the true estimate. One randomly chooses a select number of samples from the given observations. If \( \delta < 0.5 \) is the fraction of data having correspondence mismatches, and \( M \) is the number of random sampling of \( k \)-tuples of data, then, [16],

\[
P_e = 1 - P_s = 1 - \left[ 1 - \frac{\binom{N-1}{k}}{\binom{N}{k}} \right]^M
\]

\[
\approx 1 - (1 - (1 - \delta)^k)^M, \text{ for } k \ll N,
\]  

where \( P_e \) is the probability that at least one \( k \)-tuple chosen from \( M \) samples is free from outliers. The number of random sampling \( M \) required for a given amount of outlier contamination \( (\delta) \) and for a particular choice of \( P_e \) is given in Tables 1 and 2. These tables indicate that one actually has to compute only a small number of trial solutions even for a very high value of \( P_e \).

To improve the efficiency of the LMedS estimator [7], we select \( N/2 \) data points for which the residuals are the smallest using the LMedS estimate computed as above. The other half of the data may be considered to be outliers in the worst case (i.e., \( \delta = 0.5 \)). This requires the sorting \( (O(N \log N)) \) of the half of the residuals. The corresponding estimate may be called the trimmed least median of squares (TLMedS) estimate.

### 4 Simulation and Results

For the range data, an arbitrary shaped 3-D object consisting of a total of 50 feature points was generated using a random number generator. We assume that the points which could be placed anywhere on the object are visible at both instants of time (e.g., a wireframe, transparent object). A Gaussian random noise with zero mean and standard deviation 10 was added. The object was then rotated by an angle 30° around an axis passing through the origin with direction cosines \((0.6040 0.7198 0.3421)\). We added independent Gaussian random noise with zero mean and varying standard deviations \((\sigma)\) to the object points in both frames. A certain fraction \( \delta \) of the given feature points were randomly permuted to obtain a mismatch in feature point correspondence. The CLS method as described in section 2 was used to estimate the motion parameters. The results for the nonrobust CLS estimator for different values of \( \sigma \) are given in Table 3 (CASE A). Then we used a \( 4 \)-tuple of observations at a time, and obtained the LMedS estimate (CASE B). Next, results in CASE B were chosen as the initial estimate, the residuals were sorted according to their magnitude, and only the smallest half of the residuals were retained. The CLS scheme was then used on the trimmed data set to obtain the corresponding TLMedS estimate (CASE C). To show that the LMedS estimates do not depend on the magnitude of the outliers, the coordinates of feature points in the first frame in CASE A, were multiplied by a factor of 10. The corresponding CLS estimates (CASE A) do not at all resemble the actual values of the parameters. However the corresponding LMedS and the TLMedS estimates remain unchanged.

For the results in Table 3, only 50 Monte-Carlo trials have been used, and the true parameters were recovered. The amount of mismatches in point correspondence was then increased to \( \delta = 0.3 \). The results are given in Table 4. Only 200 Monte-Carlo trials were
performed, and the motion parameters were correctly recovered. For perspective projection case, we used the same data set, but shifted along the z-direction so that the depth is always positive (in both frames). The object underwent a rotational as well as a translational motion. The results are given in Tables 5 and 6. \( P_0 \) tolerated in Monte-Carlo sampling was less than 0.01%. It may be seen from the results in Tables 5 and 6 that the non-robust estimate of the essential matrix is very poor in presence of correspondence mismatches. Although the correct parameters could be extracted using the proposed LMedS scheme, the estimates are quite sensitive to the sensor noise. The reason is that the elements of the estimated \( E \) matrix are not independent [17].

5 Conclusion

Since the pre-processing step of establishing the feature (point) correspondence between successive time frames is a very difficult task, one usually achieves such correspondence with only limited accuracy. By virtue of having a very high breakdown point, the LMedS estimator provides the much desired robustness against the correspondence mismatches. The applicability and the design procedure of a robust algorithm for the processing of a long sequence of motion data are currently under investigation.

References


Table 1: Number of random sampling required for the motion estimation from range data.

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( k )</th>
<th>( P_0 = 95% )</th>
<th>( P_0 = 99% )</th>
<th>( P_0 = 99.9% )</th>
<th>( P_0 = 99.99% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>0.30</td>
<td>4</td>
<td>12</td>
<td>17</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>26</td>
<td>37</td>
<td>56</td>
<td>74</td>
</tr>
</tbody>
</table>

Table 2: Number of random sampling required for the motion estimation from perspective projection data.

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( k )</th>
<th>( P_0 = 95% )</th>
<th>( P_0 = 99% )</th>
<th>( P_0 = 99.9% )</th>
<th>( P_0 = 99.99% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>14</td>
<td>19</td>
<td>29</td>
<td>38</td>
</tr>
<tr>
<td>0.30</td>
<td>10</td>
<td>113</td>
<td>161</td>
<td>242</td>
<td>322</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>231</td>
<td>331</td>
<td>496</td>
<td>661</td>
</tr>
</tbody>
</table>
Table 3: Estimates of the motion parameters for the range data in presence correspondence mismatches. Here, A) CLS, B) LMedS, and C) TLMedS estimates.

Table 4: Estimates of the motion parameters for the range data when there are more outliers in the data.

Table 5: Estimates of the essential parameters for visual image data in presence of correspondence mismatches.

Table 6: Estimates of the essential parameters for visual image data when there are more outliers in the data.