Estimating 3-D Motion and Structure Parameters Of A Rigid Planar Object By Establishing Landmark Correspondences

Nirwan Ansari and Herman Agustiawan
Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights
Newark, New Jersey 07102
USA

Abstract – Results of marrying between a shape recognition method [1] and two motion estimation algorithms [2],[3] to perform motion and structure estimation of a rigid planar object from landmark matching are presented. From all numerical experiments that have been tried, the recognition algorithm can handle the landmark matching tasks well if the object follows a relatively small rigid motion such that the normal of the object surface is not almost perpendicular to the optical axis. Due to the nonlinearity of the computation procedure of the married algorithm and roundoff error generated by the computer, the estimated parameters deviate up to 10% from the predefined parameters. The married algorithm is, however, convenient, inexpensive and not shaken by missing or extraneous landmarks in the second view as long as there are more than three landmark corresponding pairs.

I. INTRODUCTION

In their correspondences [2],[3], Tsai, Huang and Zhu indirectly proved that at least four feature point correspondences over two views are sufficient to estimate motion and structure parameters of a moving rigid planar patch. Meanwhile, Ansari and Delp [1] have used feature points in the object, known as landmarks, to represent the entire contour of the object and perform the matching task for recognizing planar objects in 3-D space. Their matching algorithm, probe and block, is based on sphericity values derived from the affine transformations that map landmarks of a model object to those of a scene.

The algorithm proposed by Tsai et al is based on a unique mapping given four or more landmark correspondences. Their algorithm is distinguished from the object following a rigid motion with either small [2] or large [3] rotation. In this paper we shall present results of marrying both the shape recognition method [1] and the motion estimation algorithms [2],[3] to perform motion and structure estimation from landmark matching. A rather detailed review on motion estimation can be found in [4],[5].

II. ESTIMATING MOTION AND STRUCTURE FROM LANDMARK MATCHING : A MARRIED ALGORITHM

It is desired to have the correspondences between landmarks of the moving rigid object in the first view and those in the second view. After we have determined these correspondences, the reconstruction of the 3-D object along with the motion parameters are then computed by using the motion and structure estimation algorithm. In this section, [1] and [2],[3] will be briefly described, and combined to perform the above task.

A. Affine Transformation and Sphericity

Given any two points, \( P \) and \( P' \in \mathbb{R}^2 \), the affine transformation [6] is given by:

\[
P' = AP + t,
\]

where

\[
P' = \begin{bmatrix} x' \\ y' \end{bmatrix}, \quad P = \begin{bmatrix} x \\ y \end{bmatrix}, \quad t = \begin{bmatrix} m \\ n \end{bmatrix},
\]

\[
A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{and } \det(A) \neq 0.
\]

The sphericity value of the above affine transformation is defined as [7]:

\[
S = \frac{\det(A^tA)}{\left(\frac{1}{2}\text{tr}(A^tA)\right)^2}.
\]

where \( \det() \) and \( \text{tr}() \) are the determinant and the trace of a matrix. Substituting the elements of \( A \) in Equation (1) into Equation (2) yields

\[
S = \frac{t_1^2 + t_2^2 - (t_1^4 + t_2^4)}{t_1^2 + t_2^2 + t_1^4 + t_2^4}
\]

where \( t_1 = a - d, t_2 = a + d, t_3 = b - c, \) and \( t_4 = b + c. \) The coefficients of the affine transformation \( a, b, c, d, m, \) and \( n \) in Equation (1) can be determined if three noncollinear points \( P_1, P_2, P_3 \) and their transformed points \( P_1', P_2', P_3' \) are given.

B. The Probe and Block Matching Algorithm

The probe and block matching algorithm adopts the sphericity value of the affine transformation mentioned above as a shape measure to achieve recognition. First of all, a table of compatibility between landmarks in the first and second view, as shown in Table 1, is constructed. The row indices \( i \) \((i = 1, 2, 3, ..., n)\) correspond to the projected landmarks from a 3-D object to a 2-D image in the first view, while the column indices \( j \) \((j = 1, 2, 3, ..., m)\) correspond to the projected landmarks in the second view. The \((i,j)\) entry of the table is the sphericity value of the affine transformation mapping the \(i\)th landmark along with its two adjacent landmarks in the first view to the \(j\)th landmark and their two respective adjacent landmarks in the second view. Thus, the matched landmarks correspond to sequences of entries in the table that are diagonal and have sphericity values close to each other. The probe and block algorithm [1] iteratively, based on some heuristics and the sphericity values, probes (searches) for possible matching
Table 1: Compatibility Table.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.28</td>
<td>0.85</td>
<td>0.12</td>
<td>0.14</td>
<td>0.81</td>
<td>-0.06</td>
</tr>
<tr>
<td>2</td>
<td>0.23</td>
<td>0.04</td>
<td>0.99</td>
<td>-0.49</td>
<td>-0.62</td>
<td>0.61</td>
<td>-0.48</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.07</td>
<td>-0.06</td>
<td>-0.43</td>
<td>-0.59</td>
<td>0.91</td>
<td>0.24</td>
<td>0.99</td>
<td>-0.10</td>
</tr>
<tr>
<td>4</td>
<td>-0.09</td>
<td>-0.04</td>
<td>0.52</td>
<td>0.21</td>
<td>0.89</td>
<td>0.98</td>
<td>0.34</td>
<td>-0.09</td>
</tr>
<tr>
<td>5</td>
<td>-0.08</td>
<td>0.04</td>
<td>-0.64</td>
<td>0.21</td>
<td>0.69</td>
<td>0.99</td>
<td>0.23</td>
<td>-0.08</td>
</tr>
<tr>
<td>6</td>
<td>-0.08</td>
<td>-0.05</td>
<td>0.46</td>
<td>0.57</td>
<td>0.21</td>
<td>0.23</td>
<td>0.99</td>
<td>-0.10</td>
</tr>
<tr>
<td>7</td>
<td>0.07</td>
<td>0.14</td>
<td>0.52</td>
<td>-0.46</td>
<td>0.52</td>
<td>0.62</td>
<td>-0.48</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 1: Compatibility Table.

Entries and blocks (eliminates) impossible matching entries in the compatibility table from further consideration.

To verify the matched landmarks which result from the probe and block algorithm, [1] uses the overall goodness of match derived from the least squared error as follows:

$$
\varepsilon' = \left[ 1.0 + \frac{n - 2}{k - 2} \log(n - 2) \right] \varepsilon,
$$

(4)

where:

- $\varepsilon' =$ matching error,
- $n =$ total number of landmarks in the first view ($n \geq k$),
- $k =$ total number of matched landmarks ($k \geq 3$),
- $\varepsilon = \varepsilon'/(\text{scale factor}) =$ normalized squared error, $\varepsilon =$ least squared error.

Table 1 shows the compatibility table derived from two views of a moving aircraft. The aircraft is rotated about an arbitrary axis through the origin by $10^\circ$ with the direction cosine ($\cos 45^\circ, \cos 60^\circ, \cos 120^\circ$), and then translated by $(16, 20, 15)$ from the original position. Here, we are considering two extraneous landmarks in the second view, i.e., $n = 7$ and $m = 9$. The resulting possible matches of landmarks are entries (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7) and (7, 8), and the corresponding matching error is 0.18.

C. Estimating Motion and Structure of A Rigid Planar Patch

Results obtained from [2] and [3] which use the same basic geometry as shown in Figure 1 will be repeated here. It was shown that the unique mapping from the first to the second view, $(X, Y, Z)$ to $(X', Y')$, of the projected landmarks of the object is given by:

$$
X' = \frac{a_1 X + a_2 Y + a_3}{a_4 X + a_5 Y + a_6} + 1 ; \quad Y' = \frac{a_1 X + a_2 Y + a_3}{a_4 X + a_5 Y + 1}
$$

(5)

for small rotation:

$$
a_1 = (1 + a_3 t) / m
$$
$$
a_2 = (-n_3 \sin \theta + b_3 a_3 t) / m
$$
$$
a_3 = (n_3 \sin \theta + a_3 t) / F
$$
$$
a_4 = (n_3 \sin \theta + a_3 t) / F
$$
$$
a_5 = (1 + a_3 t) / m
$$
$$
a_6 = (b_3 \sin \theta + d_3 a_3 t) / F
$$
$$
a_7 = (b_3 \sin \theta + d_3 a_3 t) / F
$$
$$
a_8 = (n_3 \sin \theta + b_3 a_3 t) / m F
$$

(6)

for large rotation ($F = 1$):

$$
a_1 = [n_2^2 + (1 - n_2^2) \cos \theta + a_3 t] / k
$$
$$
a_2 = [n_2 n_3 (1 - \cos \theta) - n_2 \sin \theta + b_3 a_3 t] / k
$$

where:

$$
(a_1, a_2, a_3) = \text{translation vector}
$$
$$
(n_1, n_2, n_3) = \text{direction cosine}
$$
$$
(a, b, c) = \text{structure parameters}
$$
$$
\theta = \text{rotation angle}
$$
$$
F = \text{focal length}
$$

From Equation (4), it is clear that if four or more landmark corresponding pairs are given or obtained from the matching algorithm, the parameters $a_i$'s can be determined. When the rotation is small, the unknown parameters $\Phi_1$, $\Phi_2$, $\Delta x$, $\Delta y$, $\Delta z$, $a$, $b$, and $c$, where $\Phi_i = \pi \theta_i$, can be determined after we have solved a sixth order polynomial equation. Note that, as pointed out in [3], $\Delta z$ is a scale factor and cannot be determined. Therefore, we can hope to determine the $z_i$'s only to within a relative depth.

When the rotation is large, the motion and structure parameters can be computed after decomposing the singular values of a $3 \times 3$ real matrix consisting of the parameters $a_i$'s. The number of possible solutions depends on the multiplicity of the resulting singular values (3 cases) as follows. Case 1: Multiplicity = 2, rotation around an axis through the origin followed by a translation along the normal direction of the object surface (unique solution). Case 2: Multiplicity = 1, rotation around an axis through the object followed by a translation along a direction different from the normal direction (two solutions). Case 3: Multiplicity = 1, rotation around an axis through the origin only (unique solution). Closed form solutions to the motion and structure for both small and large rotations are given in [2], [3], respectively.

Figure 1: Basic Geometry.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

We have performed various experiments, considering cases with complete and incomplete landmark corresponding pairs.
A. Experiments with Complete Landmark Corresponding Pairs

Table 2 shows the predefined motion and structure parameters for both small and large motions, while Table 3 shows their estimated parameters. Figure 2 shows the landmarks ("•") of the object before and after following a small rigid motion, whereas Figure 3 shows the landmarks before and after following a large motion: (a) Case 1, (b) Case 2, and (c) Case 3.

From Table 3, we see that for small rigid motion all landmarks in the first view match with those in the second view, and the matching error here is 0.36. It is smaller than that which we obtain in Case 1 and Case 2 for large motion, but greater than that obtained in Case 3. The reason is that when the object undergoes a large movement such that the normal of the object surface almost exactly perpendicular to the optical axis, the least squared error is large. Since the matching error is derived from the least squared error plus a penalty term for incomplete matching, the matching error is large if the matching is incomplete. In the experiment, instead of translating the object as far as two units away from the initial position along the normal direction as in Case 1 of Table 2, it is translated three and four units away from the initial position with the same direction. When it is translated three units away from the initial position, the matching error is 4.01 and only four matched landmarks are obtained from a total of six. Therefore, a penalty is added here to the least squared error for incomplete matching. When it is translated four units away, the matching algorithm fails to detect the correct matches of landmarks, i.e., the matching error is undefined. The reason may be attributed to the fact that, because of the motion, the normal of the object surface is almost perpendicular to the view axis. In contrast, for Case 3, the matching error is 0.06 because the object follows a rotation around an arbitrary axis through the origin only (no translation).

Although the probe and block algorithm can perform the matching task well if the object follows only a small rigid motion, it does not mean that the estimated motion parameters will be entirely dependent on how well the landmarks in the first view can match to those in the second view. That is, it still depends on the algorithm that we use to estimate these unknown motion parameters. For example, in the small motion algorithm, \( \Delta z \) in Equation (6) is a scale factor and cannot be determined, and therefore the solution to the motion is a function of \( \Delta z \) which can be computed only if it is assumed as a relative depth of the object. In addition, as a result of solving

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Motion</th>
<th>Large Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation Angle</td>
<td>2°</td>
<td>10° (all cases)</td>
</tr>
<tr>
<td>Direction Cosine</td>
<td>(cos 58°, cos 122°, cos 45°)</td>
<td>the same (all cases)</td>
</tr>
<tr>
<td>Translation Vector ( N = \sqrt{a^2 + b^2 + c^2} = \sqrt{6} )</td>
<td>case 1 ( \begin{pmatrix} 2 \hat{x}, 1 \hat{y} \end{pmatrix} )</td>
<td>case 2 ( \begin{pmatrix} 3 \hat{x}, 2 \hat{y} \end{pmatrix} )</td>
</tr>
<tr>
<td>Object Parameters</td>
<td>(1.2, 1)</td>
<td>(1.2, 1)</td>
</tr>
<tr>
<td>Number of Landmarks ( N )</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2: Predefined Motion and Structure Parameters.

Figure 2: Landmarks of a jeep: (a) before and (b) after a small rigid motion. Note that the vertices are the landmarks.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Motion</th>
<th>Large Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation Angle</td>
<td>See the exact table</td>
<td>case 1: 9.39°</td>
</tr>
<tr>
<td>Direction Cosine</td>
<td>Sol. 1: ( d_1 = 0.0147 ) ( d_2 = 0.0087 ) ( d_3 = 0.0123 )</td>
<td>case 2: (0.53, 0.52, 0.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 3: (0.53, 0.53, 0.67)</td>
</tr>
<tr>
<td>Translation Vector</td>
<td>Sol. 1: ( \Delta x^* = 1.01 ) ( \Delta y^* = 4.01 )</td>
<td>case 1: ( \omega = (0.42, 0.81, 0.41) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 2: ( \omega = (4.54, 3.69, 4.29) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 3: arbitrary</td>
</tr>
<tr>
<td>Object Parameters</td>
<td>Sol. 1: (0.35, 0.45, 0.22)</td>
<td>case 1: ( \omega = (0.42, -0.42, -0.2) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 2: ( \omega = (4.54, 3.69, 4.29) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 3: arbitrary</td>
</tr>
<tr>
<td># of Matched Landmarks</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Matching Error</td>
<td>0.36</td>
<td>case 1: 0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 2: 2.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case 3: 0.06</td>
</tr>
</tbody>
</table>

Table 3: Estimated Motion and Structure Parameters.
values are equal, then there is only one possible orientation for the object surface before motion (Case 1). Finally, if the singular values are all identical, the motion consists of rotation around an axis through the origin only (Case 3). Hence the object surface can be anywhere, but the solution to the motion is only one. Therefore, the results shown in Table 3 support the claim of [3].

B. Experiments with Missing and Extraneous Landmarks

In these experiments, we use the same motion and structure parameters and the object follows only a large rigid motion (Case 2) as shown in Table 5. Figure 4 shows the landmarks of the object before and after following a rigid motion : (a) missing one point; (b) adding two extraneous points. Here we present only one missing and two extraneous landmarks from the given six landmarks. From Table 6, we see that the estimated parameters are not corrupted by the missing or extraneous landmarks in the second view. As described by Equation (7), the parameters $a_i$'s are a function of the motion and structure parameters only. These parameters are resulted from a unique mapping of landmarks in the first to second view. Therefore, regardless of which corresponding landmarks on the object being used, as long as there are more than three landmark corresponding pairs, the estimated parameters will remain relatively the same. Note that, since a penalty for incomplete matching is added in the experiment with two extraneous points, the matching error could be different for both experiments, but the motion and object parameters are still relatively the same.

Figure 4: Landmarks of a jeep: before (a) and after a large rigid motion with a missing landmark (b), and with two extraneous landmarks (c).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rot. Angle</td>
<td>Sol. 1: 1.55°</td>
</tr>
<tr>
<td></td>
<td>Sol. 2: 1°</td>
</tr>
<tr>
<td>Dir. Coine</td>
<td>Sol. 1: (0.545, 0.545, 0.864)</td>
</tr>
<tr>
<td></td>
<td>Sol. 2: (0.5, 0.5, 0.707)</td>
</tr>
<tr>
<td>Trans. vector</td>
<td>Sol. 1: (0.202, 0.202, 0.20)</td>
</tr>
<tr>
<td></td>
<td>Sol. 2: (0.015, 0.015, 0.020)</td>
</tr>
<tr>
<td>Obj. Parameters</td>
<td>Sol. 1: (1.15, 1.15, 1.15)</td>
</tr>
<tr>
<td></td>
<td>Sol. 2: (1.15, 1.15, 1.15)</td>
</tr>
</tbody>
</table>

Table 4: Solution to the Small Rigid Motion.
Table 5: Predefined Motion and Structure Parameters (Further Experiments).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Missing Landmarks</th>
<th>Extraneous Landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rot. Angle</td>
<td>12°</td>
<td>ditto</td>
</tr>
<tr>
<td>Dir. Cosine</td>
<td>((\cos 45^\circ, \cos 60^\circ, \cos 120^\circ))</td>
<td>ditto</td>
</tr>
<tr>
<td>Obj. Parameters</td>
<td>((1.2, 1))</td>
<td>ditto</td>
</tr>
<tr>
<td>Trans. Vector</td>
<td>((\vec{a}, \vec{b}, \vec{c}))</td>
<td>ditto</td>
</tr>
<tr>
<td># of Lmarks @ (T_1)</td>
<td>6</td>
<td>ditto</td>
</tr>
<tr>
<td># of Lmarks @ (T_2)</td>
<td>5</td>
<td>ditto</td>
</tr>
</tbody>
</table>

Table 6: Estimated Motion and Structure Parameters (Further Experiments).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Missing Landmarks</th>
<th>Extraneous Landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rot. Angle</td>
<td>11.12°</td>
<td>11.06°</td>
</tr>
<tr>
<td>Dir. Cosine</td>
<td>((0.71,0.52,-0.32))</td>
<td>((0.69,0.48,-0.48))</td>
</tr>
<tr>
<td>Obj. Parameters</td>
<td>((0.95,0.05,3.73))</td>
<td>((0.95,0.76,3.55))</td>
</tr>
<tr>
<td>Trans. Vector</td>
<td>((\vec{a}, \vec{b}, \vec{c}))</td>
<td>((\vec{a}, \vec{b}, \vec{c}))</td>
</tr>
<tr>
<td># of Matched Lmarks</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Matching Error</td>
<td>2.67</td>
<td>3.49</td>
</tr>
</tbody>
</table>

IV. CONCLUSIONS

- The probe and block algorithm performs the landmark matching task well if the object follows a motion such that the normal of the object surface is not almost perpendicular to the optical axis. The algorithm, without too much destroying the sequential order of the original landmarks, is also capable of detecting the correct matches of landmarks when missing and some extraneous landmarks in the second view are taken into account.

- When the rigid motion is small, only a relative depth of the object can be obtained. The number of solutions, aside from the scale factor for the translation, depends on the number of the real roots of the sixth order polynomial equations. If the motion is large, the number of solutions is either one or two depending on the multiplicity of the singular values.

- The motion estimation algorithm is convenient, inexpensive and not shaken by missing or extraneous landmarks in the second view as long as there are more than three landmark corresponding pairs. That is, regardless of which landmarks in the object being used, the estimated motion and structure parameters will remain relatively the same.

V. REFERENCES


