TWO VIEW MOTION ANALYSIS, STEREO VISION AND A MOVING CAMERA'S POSITIONING
THEIR EQUIVALENCE AND A NEW SOLUTION PROCEDURE

Xinhua Zhuang*, Robert M. Haralick, Fellow IEEE

Machine Vision International
Ann Arbor, MI 48104

ABSTRACT

The three problems: Two view motion analysis, stereo vision and determining a moving camera's position are all the same problem. We explain their equivalence and introduce a noise robust procedure for solving these problems.

1. INTRODUCTION

Research on two view motion analysis, stereo vision and determining a moving camera's position has proceeded mostly separately. Two view motion analysis ([1] [2]) is concerned about determining the motion and the visible surface structure of a moving object from two pictures taken at two times by a camera. Stereo vision [3] is concerned about determining the visible surface structure of a stationary object from two pictures taken at two places. A moving camera's position problem [5] is concerned about determining the change of the moving camera's position from two pictures taken by the camera at two successive times. Researchers in computer and robot vision circles have realized that there exist some interesting connections between these problems. Exploring the connections is meaningful since some significant results in one field can be more or less transplanted into other fields. Sometimes, the appropriateness of the representation in one field makes things easier to do than in another field.

This article shows that all three problems: two view motion analysis, stereo vision and determining a moving camera's position are all the same problem. Also a procedure which, the authors believe is adjustable to the noise case for solving the two view motion analysis problem is described.

In section 2, we review some basic mathematics of frame transformation which we will use in all the remaining sections of the paper.

Section 3 stipulates the relationship between a camera's position and some suitable frame.

*Visiting Scientist from Zhejiang Computing Technique Institute, Hangzhou, People's Republic of China

Section 4 introduces the surface assumption and two view motion analysis results which are obtained recently by authors.

Section 5 identifies the stereo vision problem with the relative depth problem of two view motion analysis.

Section 6 identifies the problem of determining a moving camera's position with determining the rigid body motion in two view motion analysis.

The final section is a conclusion.

2. COORDINATE FRAME TRANSFORMATION AND RIGID BODY MOTION

In this section, we show the relationship between coordinate frame transformation and rigid body motion.

Suppose that there are two right hand coordinate frames in the 3-D Euclidian space $E_3$: $F = \{e_1, e_2, e_3\}$ and $F' = \{e_1', e_2', e_3'\}$. The orthonormal basis $(e_1, e_2, e_3)$ can be represented in the orthonormal basis $(e_1', e_2', e_3')$ by means of an orthonormal matrix $R_O$ of the first kind (i.e. $\det (R_O) = 1$) as follows

$(e_1, e_2, e_3) = (e_1', e_2', e_3') R_O$

Denote the translation from the origin $o'$ to the origin $o$ by the vector $T_O$. Then,

$o = o' + T_O$

Any point $p$ in $E_3$ can be represented in both $F$ and $F'$ as follows:

$p = xe_1 + ye_2 + ze_3 = x'e_1' + y'e_2' + z'e_3' + o$

Thus,

$(e_1', e_2', e_3') (x', y', z')^t$

$= (e_1, e_2, e_3) (x, y, z)^t + T_O$

$= (e_1', e_2', e_3') R_O (x, y, z)^t + T_O$

where "$t$" represents the transposition operation.
Suppose that the frame $F'$ coincides with the standard frame in $E_3$, that is,
\[
\begin{align*}
o' &= (0,0,0)t \\
e'_1 &= (1,0,0)t \\
e'_2 &= (0,1,0)t \\
e'_3 &= (0,0,1)t
\end{align*}
\]
then there holds
\[(x'y'z')^t = R_{O} (x,y,z)^t + T_{O}
\]
In other words, when the standard frame $F'$ in $E_3$ is transformed into the right hand frame $F$ by a rotation $R_{O}$ and the translation $T_{O}$, the point $(x,y,z)$ experiences a rigid body motion defined by $R_{O}, \ R_{O}$ and becomes the point $(x',y',z')$ where $(x,y,z)$ and $(x',y',z')$ are the coordinate representations of the same spatial point $p$ in $E_3$ with respect to the frames $F$ and $F'$ respectively.

The rigid body motion could be represented more concisely by the homogeneous transformation:
\[
H = \begin{bmatrix}
R_{O} & T_{O} \\
0 & 1
\end{bmatrix}
\]
Conversely, assume a rigid body motion $(T_{O},\ R_{O})$ given which moves a point $p$ into another point $p'$ in the space $E_3$. Suppose that $p$ and $p'$ have the coordinates $(x,y,z)$ and $(x',y',z')$ respectively in the coordinate frame $F$. Thus,
\[
(x',y',z')^t = R_{O}(x,y,z)^t + T_{O}
\]
Define a new frame $F' = \{o', e'_1, e'_2, e'_3\}$ as follows:
\[
(e'_1, e'_2, e'_3) = (e_1, e_2, e_3)R_{O}^t
\]
and
\[
o' = o - T_{O}
\]
Consider $F'$ as the standard frame. In that case, $e'_1, e'_2, e'_3$ and $x', y', z'$ represent the same point in $E_3$ and hence $(x',y,z)$ and $(x',y',z')$ become coordinate representations of that point with respect to $F$ and $F'$ respectively.

### FRAME AND CAMERA. IMAGE PLANE.

### PERSPECTIVE PROJECTION

In this section we stipulate a camera's position with respect to an appropriate frame in such a way that its position (camera's lens and orientation) is identified with the frame. There are three cases:

1. **Two view motion analysis:** A camera is sed. The camera's lens coincides with the frame's origin 0, its view line directs z0, and its image planes are z=f and z=f' at the two successive times t and t' respectively where f and f' are focal lengths of the camera at times t and t' respectively. The point $(x,y,z)^t$ is moved into the point $(x',y',z')^t$ by the rigid body motion $(T_{O},R_{O})$ from the time t to t' and p, p' are projected onto the image planes $z=f$ and $z=f'$ respectively and become
\[
(X,Y,I) = (xI/z,yI/z,1)
\]
and
\[
(X',Y',I') = (x'I'/z',y'I'/z',1')
\]
respectively.

Since the focal lengths f and f' are assumed known, without loss of generality, they could be normalized out as 1. After normalization the perspective projections $(X,Y)$ and $(X',Y')$ are called an image point correspondence pair.

2. **Stereo vision:** two cameras are used. One's lens coincides with frame F's origin o and its view line directs towards zo. Another's lens coincides with the frame F's origin o' and its view line directs towards z'0.

Suppose that z=f and z'=f' are two image planes in F and F' respectively. The same spatial point p is projected onto the image planes $z=f$ and $z'=f'$ respectively by a central perspective projection. Without loss of generality, the focal lengths f and f' are assumed 1 and the perspective projections $(X,Y)$ and $(X',Y')$ are called an image point correspondence pair.

3. **A moving camera's position:** A moving camera is used. The camera before motion is identified with F and after motion is identified with F'. The camera's lens coincides with the origin of the corresponding frame. The camera's view line directs towards z0 or z'0 before or after motion respectively. Normalizing the focal lengths, we call the perspective projections $(X,Y)$ and $(X',Y')$ as an image point correspondence pair.

Given a set of image point correspondence motion $(T_{O},R_{O})$ in case 1 to determine the visible pairs, is it possible to determine the rigid body motion $(T_{O},R_{O})$ in case 1, to determine the visible surface structure in case 2 and to determine the change of the moving camera's position in case 3?

4. **SURFACE ASSUMPTION. TWO VIEW MOTION ANALYS**

Suppose that the image point correspondence pair set P comes from a surface patch or a group of surface points S in $E_3$. 

7
It is proved ([1]) that the mode of motion (whether $T_0$ is zero or not), the rotation $R_0$, the translational orientation $T_0/||T_0||$ (when $T_0 \neq 0$) where $||.||$ represents the Euclidian vector norm and the relative depth $z/z'$ all are determined unambiguously by the correspondence pair set $P$ if and only if the surface assumption holds, that is, the surface patch or the group of surface points $S$ cannot be contained in a quadratic surface of form.

$$(x,y,z) U (x,y,z)^T + v^T (x,y,z)^T = 0$$

with $||U-U_0|| + ||V|| \neq 0$ and $T_0^T R_0 U = v^T$

where "$t$" represents the transposition operation. A sufficient condition to guarantee all these nice things is: the surface patch or the group of surface points $S$ can not be contained in a quadratic surface which passes through the origin $0$.

In the following we always assume that the surface assumption holds.

Let

$$A = (X,X',Y,Y',X',Y')$$

$$W = \sum_{P} \alpha_{P} A^T A$$

$$h = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9)$$

$$E = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$$

where $[(X,Y), (X',Y')] \in P$, an image point correspondence pair.

Under the surface assumption it is proved (see([1])) that there are 8 linear independent $A_i$'s denoted by $A_1, \ldots, A_8$ such that $\text{Rank}(W) = \text{Rank}(\sum \alpha_i A_i^T A_i) = 8$ and hence the general solution $h$ is one parameter if and only if $T_0 \neq 0$ and there are 6 linear independent $A_i$'s denoted by $A_1, \ldots, A_6$ such that $\text{Rank}(W) = \text{Rank}(\sum \alpha_i A_i^T A_i) = 6$ and hence the general solution has three-parameters if and only if $T_0 = 0$.

As a result at least 6(8) points must be contained in $P$ depending on whether or not $T_0$ is zero. More image-point correspondence pairs are preferable to guarantee that the surface assumption holds and to smooth out any noise effects. The general solution holds the same, $E$ of the linear equation, has two and only two decompositions:

$$E = T_x R = (-T) X R'$$

where both $R$ and $R'$ are orthonormal matrices of the first kind. One of them equals $R_0$, and depending on whether $T_0$ is zero or not $T$ is any real vector or equals $\alpha T_0$ with $\alpha$ any real number.

(Note: we use "$x" to represent the cross product of two vectors. If, for instance, $r_1,r_2,r_3$ are column vectors of $R$, then

$$T_x = [T_x r_1, T_x r_2, T_x r_3]$$

It is proved in [1] that $T_0 = 0$ and $R_0 = R$ hold iff

$$1/N \sum ||v' || W ||V|| - R V ||V|| ||V|| = 0$$

where $v = (x,y,z)^T$, $v' = (x',y',z')^T$, $N = |P|$, the number of image point correspondence pairs.

When $T_0$ is not zero, we have that

$$R_0 = R \text{ and } T_0/||T_0|| = T/||T|| \text{ iff}$$

$$1/N \sum ||T_x V|| W' - T_x V ||V|| R V - T/||T|| ||T|| ||V|| = 0$$

and

$$R_0 = R \text{ and } T_0/||T_0|| = -T/||T|| \text{ iff}$$

$$1/N \sum ||T_x V|| W - T_x V ||V|| R V + T/||T|| ||T|| ||V|| = 0$$

Finally, the relative depth $z/z'$ equals $W ||V||/||V||$ or $||T_x R_0 V||/||T_x V||$ depending on whether or not $T_0$ is zero.

5. STEREO VISION

Assume that that standard frame $F'$ is transformed into $F$ by $(T_0, R_0)$ and the translation $T_0$ which moves the origin $O'$ of the frame $F'$ into the origin $O$ of the frame $F$ is not zero.

Given the image point correspondence pair set $P$, the problem of how to determine the relative surface structure (i.e., the coordinate ratio $z/z'$) is called "stereo vision". The problem could be transformed into a two view motions analysis problem. Because of section 2, we can consider $(x,y,z)^T$ and $(x',y',z')^T$ as the coordinate representations of a spatial point under the rigid body motion $(T_0, R_0)$ with respect to the frame $F$ and identify the relative surface structure problem in stereo vision with the relative depth problem under the rigid body motion $(T_0, R_0)$. Using the results in the previous section, the procedure for obtaining the depth information is as follows:

Step 1. Solve $\min h^T W h$ with $||h|| = 1$

Step 2. Let

$$L_1 = [h_1, h_2, h_3]$$

$$L_2 = [h_4, h_5, h_6]$$

$$L_3 = [h_7, h_8, h_9]$$

$T_x R = [T_x r_1, T_x r_2, T_x r_3]$.
\[ E = \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix} \]

**Step 3.** Let
\[
\alpha = (||L_2||^2 + ||L_3||^2 - ||L_1||^2)/2 \\
\beta = (||L_3||^2 + ||L_1||^2 - ||L_2||^2)/2 \\
\gamma = (||L_1||^2 + ||L_2||^2 - ||L_3||^2)/2
\]

**Step 4.** If \(|\alpha| > |\beta|, |\gamma|\), then let
\[
T = \begin{bmatrix} \sqrt{\alpha} \\ -\langle L_1, L_2 \rangle / \sqrt{\alpha} \\ -\langle L_1, L_3 \rangle / \sqrt{\alpha} \end{bmatrix}
\]
\[
r_1 = (L_2 \times L_3) \times L_1 + \sqrt{\alpha} (L_2 \times L_3) / (||T||^2 \sqrt{\alpha}) \\
r_1' = (L_2 \times L_3) \times L_1 - \sqrt{\alpha} (L_2 \times L_3) / (||T||^2 \sqrt{\alpha}) \\
r_2 = (r_1 \times L_2) / \sqrt{\alpha} \\
r_2' = -(r_1 \times L_2) / \sqrt{\alpha} \\
r_3 = (r_1 \times L_3) / \sqrt{\alpha} \\
r_3' = -(r_1 \times L_3) / \sqrt{\alpha}
\]

**GO TO STEP 7**

(Where \(\langle ., . \rangle\) represents the scalar product between two row vectors. For convenience the cross product operation 'x' also acts on two row vectors and produces a row vector.)

**Step 5.** If \(|\beta| > |\gamma|\), then let
\[
T = \begin{bmatrix} -\langle L_1, L_2 \rangle / \sqrt{\beta} \\ \sqrt{\beta} \\ -\langle L_1, L_3 \rangle / \sqrt{\beta} \end{bmatrix}
\]
\[
r_2 = (L_3 \times L_1) \times L_2 + \sqrt{\beta} (L_3 \times L_1) / (||T||^2 \sqrt{\beta}) \\
r_2' = (L_3 \times L_1) \times L_2 - \sqrt{\beta} (L_3 \times L_1) / (||T||^2 \sqrt{\beta}) \\
r_3 = (r_2 \times L_3) / \sqrt{\beta} \\
r_3' = -(r_2 \times L_3) / \sqrt{\beta} \\
r_1 = (r_2 \times L_1) / \sqrt{\beta} \\
r_1' = -(r_2 \times L_1) / \sqrt{\beta}
\]

**GO TO STEP 7**

**Step 6.** Let
\[
T = \begin{bmatrix} -\langle L_1, L_2 \rangle / \sqrt{\beta} \\ \sqrt{\beta} \\ -\langle L_1, L_3 \rangle / \sqrt{\beta} \end{bmatrix}
\]
\[
r_3 = (L_1 \times L_2) \times L_3 + \sqrt{\beta} (L_1 \times L_2) / (||T||^2 \sqrt{\beta}) \\
r_3' = (L_1 \times L_2) \times L_3 - \sqrt{\beta} (L_1 \times L_2) / (||T||^2 \sqrt{\beta})
\]
\[
r_1 = (r_3 \times L_1) / \sqrt{\beta} \\
r_1' = -(r_3 \times L_1) / \sqrt{\beta} \\
r_2 = (r_3 \times L_2) / \sqrt{\beta} \\
r_2' = -(r_3 \times L_2) / \sqrt{\beta}
\]

**Step 7.** Let
\[
R = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \\
R' = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}
\]

**Step 8.** If \(\frac{1}{N} \sum \| Tx \times R \| V - \| Tx \times V \times R \| V \| T \| = 0\), then the relative depth \(z' / z\) is given by
\[
z' / z = \| Tx \times V \| / \| Tx \times V \|
\]
(where \(V=(X,Y,1)^t\) and \(V'=(X',Y',1)^t\). \(N\) is the number of pairs in \(P\).)

**Step 9.** Otherwise, it must happen that
\[
\frac{1}{N} \sum \| Tx \times R \| V - \| Tx \times V \times R \| V \| T \| \geq \| V' \times R \| T \| (\| T \| \times V \| V \|) = 0
\]
and hence the relative depth \(z' / z\) is given by
\[
z' / z = \| Tx \times V \| / \| Tx \times V \|
\]

**Step 10.** **STOP**

When the translation \(T_0\) is not zero and given, the absolute depth \(z\) and \(z'\) could be uniquely determined as follows
\[
z' = -\| T_0 \times R_0 \| V / \| V' \| / R_0 \| V \|
\]
\[
z = -\| T_0 \times V \| / \| V' \| / R_0 \| V \|
\]

**6. CAMERA'S POSITION**

A camera's position can be completely characterized by a suitable frame. Suppose that the camera moves from a position \(Pos\) to another position \(Pos'\). Corresponding frames associated with \(Pos\) and \(Pos'\) are denoted by \(F\) and \(F'\) respectively. The homogeneous transformation \(H\) from the frame \(F'\) to the frame \(F\) is still denoted by
H = \[
\begin{bmatrix}
R_O & T_O \\
\hline
0 & 0 \\
0 & 0 \\
0 & 1 \\
\end{bmatrix}
\]

(see section 2). The relationship between Pos and F or Pos' and F' is related again as follows:

a. the camera's lens always coincides with the origin of the associated frame.

b. the image plane in Pos (or Pos') is z=f (or z'=f') where f (or f') is the focal length of the camera in Pos (or Pos'). Without loss of generality, f and f' are assumed 1.

c. the change of the camera's orientation between two frames from F to F' is described by

R_{0}=\text{rot}(\uphi) \text{ rot}(\psi) \text{ rot}(\theta)

where \text{rot}(\theta) represents a rotation of angle \theta around the y-axis in the frame F, \text{rot}(\phi) represents a rotation of angle \phi around the new x-axis and \text{rot}(\psi) represents a rotation of angle \psi around the new z-axis. Angles \theta, \phi, \psi are called pan, tilt and swing angle respectively and

\text{rot}(\theta) = \begin{bmatrix}
\cos\theta & 0 & \sin\theta \\
0 & 1 & 0 \\
-\sin\theta & 0 & \cos\theta
\end{bmatrix}, -\pi \leq \theta \leq \pi

\text{rot}(\phi) = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos\phi & \sin\phi \\
0 & -\sin\phi & \cos\phi
\end{bmatrix}, -\pi/2 \leq \phi \leq \pi/2

\text{rot}(\psi) = \begin{bmatrix}
\cos\psi & \sin\psi & 0 \\
-\sin\psi & \cos\psi & 0 \\
0 & 0 & 1
\end{bmatrix}, 0 \leq \psi \leq 2\pi

7. CONCLUSION

Two view motion analysis, stereo vision and a moving camera's position, are three identical problems. We believe that some new results on two view motion analysis (see [1] and section 4 in the article) are important and easily adjustable to the noisy case. That is why we explain in the article that all three problems are equivalent and introduce a noise robust procedures for two view motion analysis into problems of stereo vision and a moving camera's positioning.

Finally, we have to mention here that the listing of references is by no means complete since the main concern of the article is about the equivalence within three different problems. Anyhow, the readers being interested in any of these problems can find more information and more references in [1]-[4].

8. REFERENCE


