A NEW ROBUST 3D MOTION ESTIMATION UNDER PERSPECTIVE PROJECTION

*Hye Ri Cho *, Kyoung Mu Lee**, and Sang Uk Lee****

*LG Electronics Inc. Advanced Telecom. Research Lab., Korea
** the Dept. of Electrical and Electronics Engineering, Hong-Ik University, Korea
***the Dept. of Electrical Engineering, Seoul Nat'l University, Korea
(E-mail: hyeri@lge.com kmlee@wow.hongik.ac.kr sanguk@sting.snu.ac.kr).

ABSTRACT

In this paper, we present a new 3D camera motion estimation technique using optical flow from a pair of images taken under a perspective projection. The problem formulation leads to the solution of overdetermined nonlinear system of equations w.r.t. the motion parameters. By employing an efficient initial guess algorithm which uses a weak perspective projection and an image coordinate normalization technique, the nonlinear solution can be obtained robustly and accurately. The proposed method has been tested on both several synthetic and real image sequences. The results show that the performance of the proposed algorithm is quite superior to the conventional ones even under more general and noisy situations.

1. INTRODUCTION

Since 2D motion between sequence of images can be exploited to infer the relative 3D motion information between a scene and a camera, many techniques have been developed to estimate 3D motion parameters from 2D motion field in images so far [1]. In general, the problem formulation leads to the solution of overdetermined nonlinear system of equations w.r.t. the motion parameters, and usually direct nonlinear minimization approaches suffer from the initial value problem. Thus, several researches have tried to approximate and formulate the original nonlinear problem as a linear one so that the solution can be obtained easily [2, 3]. However, due to very restrictive assumptions on the motion and camera geometry, the results are not reliable in practice.

In this paper, we propose a new 3D motion parameter recovery algorithm between a camera and objects in a scene, which is formulated by nonlinear equations under more general situation without additional assumptions. In order to solve the nonlinear equations robustly, we propose an elegant initial guess algorithm, in which a weak perspective projection and an image coordinate normalization technique are employed to decouple the translational parameters from rotational ones, so that a robust initial value for the solution of the original nonlinear problem can be obtained easily. By using the estimated initial value, the motion parameters can be recovered accurately through a nonlinear minimization technique. The overall procedure of the proposed 3D motion parameter estimation algorithm is illustrated in Fig. 1.

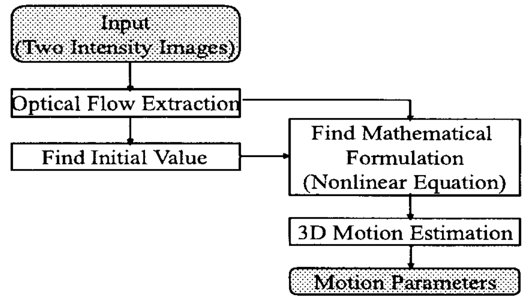

Fig. 1. Block diagram of proposed 3D motion estimation.

2. 3D MOTION ESTIMATION

2.1. Camera Model

Although no distinction is made between the situations where a) a camera is moving and objects are stationary, b) a camera is stationary while objects are in motion, or c) both camera and objects are in motion, in this paper, a mathematical formulation based on the situation a) is employed. Let us denote the space points be $\mathbf{X_i} = [X_i, Y_i, Z_i]^T$ $(i = 1, \cdots, N)$ and the projected points on the reference image plane by a perspective transformation be $\mathbf{m_{0i}} = [\mathbf{x_{0i}}^T, 1]^T = [x_{0i}, y_{0i}, 1]^T$. When the camera undergoes a rotational and translational motion described by a matrix $\mathbf{R}$ and a vector $\mathbf{t}$ respectively, the corresponding point $\mathbf{x_{1i}} = (x_{1i}, y_{1i})^T$ on the next image can be described by

$$s_{1i}[x_{1i}, y_{1i}, 1]^T = s_{1i}[\mathbf{x}_{1i}^T, 1]^T = \mathbf{P}[\mathbf{R}(s_{0i}\mathbf{P}^{-1}\mathbf{m}_{0i}) + \mathbf{t}], \quad (1)$$

where s_{ki} $(k = 0, 1, \text{and}\, i = 1, \cdots, N)$ is the distance(depth) from the image plane to the object point, and

$$\mathbf{P} = \begin{bmatrix} fS_x & 0 & 0 \\ 0 & fS_y & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathbf{R} = \begin{bmatrix} \mathbf{r}_1{}^T \\ \mathbf{r}_2{}^T \\ \mathbf{r}_3{}^T \end{bmatrix}, \ \mathbf{t} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T.$$

where f is the focal length, S_x and S_y are the scaling factors, $\mathbf{r_i} = [r_{i1}, r_{i2}, r_{i3}]^T, i = 1,2,3$ are the rotation parameters, and t_x, t_y, and t_z are the translation parameters, respectively. Note that since the 3D rotation can be decomposed into three consecutive rotations around the coordinate axes, X, Y, and Z by angles α, β, and γ respectively [4], the rotation parameters can be written by

$$
\begin{aligned}
r_{11} &= \cos\beta \cos\gamma \\
r_{12} &= -\cos\beta \sin\gamma \\
r_{13} &= \sin\beta \\
r_{21} &= \sin\alpha \sin\beta \cos\gamma + \cos\alpha \sin\gamma \\
r_{22} &= -\sin\alpha \sin\beta \sin\gamma + \cos\alpha \cos\gamma \\
r_{23} &= -\sin\alpha \cos\beta \\
r_{31} &= -\cos\alpha \sin\beta \cos\gamma + \sin\alpha \sin\gamma \\
r_{32} &= \cos\alpha \sin\beta \sin\gamma + \sin\alpha \cos\gamma \\
r_{33} &= \sin\alpha \cos\gamma
\end{aligned} \tag{2}
$$

2.2. 3D Nonlinear Motion Equation

By using (1), the set of 2D displacement vectors $[u_i, v_i] = [x_{1i} - x_{0i}, y_{1i} - y_{0i}]$ for $i = 1, \cdots, N$ can be written by

$$u_i = \frac{x_{0i}r_{11} + S_x y_{0i} r_{12}/S_y + f S_x r_{13} + f S_x t_x/s_{0i}}{x_{0i} r_{31}/f S_x + y_{0i} r_{32}/f S_y + r_{33} + t_z/s_{0i}} - x_{0i} \tag{3}$$

$$v_i = \frac{S_x x_{0i} r_{22}/S_y + y_{0i} r_{22} + f S_y r_{23} + f S_y t_y/s_{0i}}{x_{0i} r_{31}/f S_x + y_{0i} r_{32}/f S_y + r_{33} + t_z/s_{0i}} - y_{0i} \tag{4}$$

Let $\mathbf{v} = [t_x, t_y, t_z, \alpha, \beta, \gamma]^T$. Then, by eliminating s_{0i} from (3) and (4), we can obtain following nonlinear equation for each point $\mathbf{x_{0i}}, i = 1, \cdots, N$:

$$
\begin{aligned}
&f_i(\mathbf{v}) = f_i(t_x, t_y, t_z, \alpha, \beta, \gamma) = \\
&t_x(fS_y x_{0i} r_{21} + fS_x y_{0i} r_{22} + f^2 S_x S_y r_{23} - x_{0i} y_{0i} r_{31} - S_x y_{0i}^2 r_{32}/S_y \\
&- f S_x y_{0i} r_{33} - v_i x_{0i} r_{31} - S_x v_i y_{0i} r_{32}/S_y - f S_x v_i r_{33}) \\
&t_y(fS_y x_{0i} r_{11} + fS_x y_{0i} r_{12} + f^2 S_x S_y r_{13} - S_x x_{0i}^2 r_{31}/S_y - x_{0i} y_{0i} r_{32} \\
&- f S_y x_{0i} r_{33} - S_y u_i x_{0i} r_{31}/S_y - u_i x_{0i} r_{32} - f S_x v_i r_{33}) \\
&t_z(v_i x_{0i} r_{11} + S_x v_i y_{0i} r_{12}/S_y + f S_x v_i r_{13} + x_{0i} y_{0i} r_{11} \\
&- S_x {y_{0i}}^2 r_{12}/S_y + f S_x y_{0i} r_{13} - S_y u_i x_{0i} r_{21}/S_x - u_i y_{0i} r_{22} \\
&- f S_y u_i r_{23} - S_y {x_{0i}}^2 r_{21}/S_x - x_{0i} y_{0i} r_{22} - f S_y x_{0i} r_{23}) = 0
\end{aligned} \tag{5}
$$

Note that the solution of (5) is the form $[\lambda t_x, \lambda t_y, \lambda t_z, \alpha, \beta, \gamma]$ indicating that there is an ambiguity with respect to the scale of the translational components.

2.3. Initial Guess Algorithm

Although the system of nonlinear equations in (5) can be solved by an iterative nonlinear solver, the optimal solution cannot be guaranteed due to the initial value problem. Thus, in this work, an accurate and robust initial guess algorithm based on a weak perspective projection model and image coordinates normalization technique is proposed, so that an accurate and stable actual 3D motion parameters can be recovered.

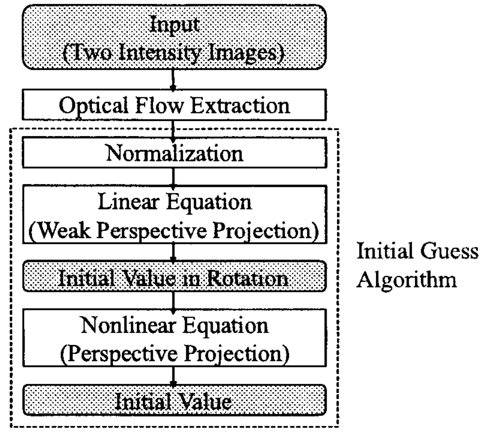

Fig. 2. Block diagram of the proposed initial guess algorithm.

Fig. 2 shows the overall flow of the proposed initial guess algorithm. The first step involves the optical flow extraction for feature points from the two consecutive images. Then, the coordinates of the feature points in each image are normalized as follows.

$$\tilde{x}_{ji} = x_{ji} - x_{jc}, \quad \tilde{y}_{ji} = y_{ji} - y_{jc}. \tag{6}$$

where x_{ji}, y_{ji} are the i-th image points $(i = 1, \cdots, N)$ in the (j+1)-th image $(j = 0, 1)$, and x_{jc}, y_{jc} are the mean (centroid) of the feature points x_{ji} and y_{ji}, respectively.

In the second step, these normalized feature points are projected onto each image plane through a weak perspective projection such that

$$\hat{s}_1 [\tilde{x}_{1i} \;\; \tilde{y}_{1i} \;\; 1]^T = \tilde{\mathbf{R}} \hat{s}_0 [\tilde{x}_{0i} \;\; \tilde{y}_{0i} \;\; 1]^T + \tilde{\mathbf{t}}, \tag{7}$$

where $\hat{s}_i (i = 0, 1)$ is the average depth in $(i + 1)$-th image, and the rotation matrix and the translation vector are defined by

$$
\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \tilde{r}_{13} \\ \tilde{r}_{21} & \tilde{r}_{22} & \tilde{r}_{23} \\ \tilde{r}_{31} & \tilde{r}_{32} & \tilde{r}_{33} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & f r_{13} \\ r_{21} & r_{22} & f r_{23} \\ r_{31}/f & r_{32}/f & r_{33} \end{bmatrix}
\cong \begin{bmatrix} 1 & -\gamma & f r_{13} \\ \gamma & 1 & f r_{23} \\ r_{31}/f & r_{32}/f & 1 \end{bmatrix}, \quad \tilde{\mathbf{t}} = \begin{bmatrix} \tilde{t}_x \\ \tilde{t}_y \\ \tilde{t}_z \end{bmatrix} \tag{8}
$$

From (8), we note that the rotational parameters along an axis orthogonal to the image plane (Z-axis) in the weak per-

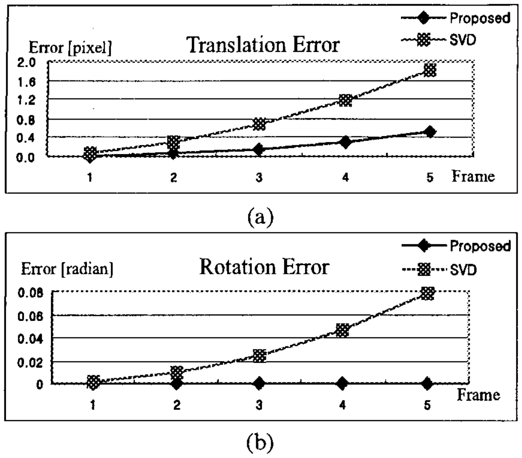

Fig. 3. Estimation accuracy. (a) Translation Error, (b) Rotation Error.

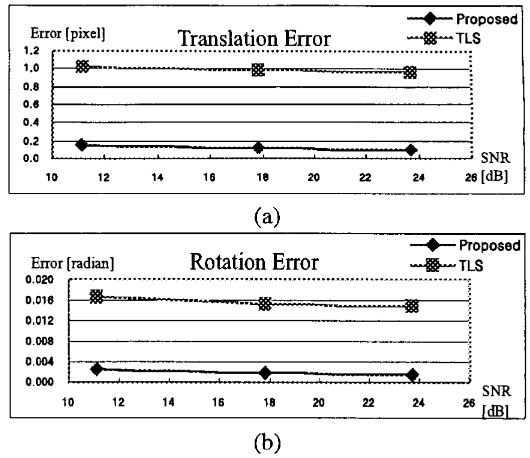

Fig. 4. The robustness of noise. (a) Translation Error, (b) Rotation Error.

spective projection is similar to those in the perspective projection, so that we can use them as the initial values for true rotation. And by corresponding the origins of the two normalized coordinates in (8), additional relationships between the rotational parameters and translational ones are obtained. Thus, by eliminating $\hat{s}_0$ from the normalized optical flow and replacing the translational parameters by the rotational ones, we can obtain a linear equation written by

$$\tilde{v}\,\tilde{x}_{0i}\tilde{r}_{11} + \tilde{v}\tilde{y}_{0i}\tilde{r}_{12} + \tilde{x}_{0i}\tilde{y}_{0i}\tilde{r}_{11} + \tilde{y}_{0i}^2\tilde{r}_{12} - \tilde{u}\,\tilde{x}_{0i}\tilde{r}_{21} + \tilde{u}\tilde{y}_{0i}\tilde{r}_{22} + \tilde{x}_{0i}^2\tilde{r}_{21} + \tilde{x}_{0i}\tilde{y}_{0i}\tilde{r}_{22} = 0 \qquad (9)$$

Once γ (r_{21}, or -r_{12}) is determined from (9), α and β can also be estimated by (2). Then, by solving (5) using these values, while fixing γ and $t_z = 1$, we can get the complete initial value.

2.4. 3D Motion Estimation

Now, let us formulate the solution of the system of nonlinear equations in (5) as the following cost minimization problem.

$$\underset{\mathrm{v}\,\in\,R^6}{\arg\min}\ \{E = \|\sum_{i=i}^{N} f_i(\mathbf{v})\|^2\} \qquad (10)$$

Then, the optimal solution can be obtained by Newton-Rhapson method [5] effectively with the initial value obtained in Section 2.3. The procedure of the proposed algorithm is summarized in the following:

1. Set the error range $\triangle E$, estimate the initial values $\mathbf{v_0}$=$[t_{x0}, t_{y0}, \alpha_0, \beta_0, \gamma_0]^T$ and fix t_z.
2. Compute $\mathbf{v_{n+1}}$=$\mathbf{v_n}$+$\triangle\mathbf{v}$ using $\triangle\mathbf{v}$, modified Newton step, partial derivative of nonlinear equation (5), optical flow u_i, v_i, and feature points x_{0i}, y_{0i} ($i = 1, \cdots, N$).
3. Calculate the evaluation function E (equation (10)) using $\mathbf{v_{n+1}}$, optical flow u_i,v_i and pixel values x_{0i}, y_{0i} (i=$1, \cdots, N$).
4. Stop if $E < \triangle E$, otherwise go to step (2) with $n = n + 1$.

3. EXPERIMENTAL RESULTS

In this section, we have tested the proposed algorithm on both synthetic and the real image sequences, and compared the results with those of conventional methods [2, 6, 7].

3.1. Experiment with Synthetic Data

For this test, we have selected 51 arbitrary 3D test points and corresponding 2D image points with known camera motion parameters. The image size is 256×256 and the focal length is set to be 477.70. The performance of the proposed 3D motion estimation has been evaluated on both noise-free and noisy images with the quantization error as well as the measurement error. First, in noise-free case, although the translation and rotation become large, the proposed algorithm estimated the motion parameters accurately. Fig. 3 shows the L2-norm error of the translation parameters [pixel] and the rotation angles [radian], which shows that the estimation accuracy of the proposed algorithm is superior to that of SVD-based method [6]. Next, Fig. 4 shows the performance of the proposed method compared to that of the TLS method [7] in noisy environments. White Gaussian measurement noise and several different levels of quantization error are assumed as in [7]. The results demonstrate that the proposed method is more robust to noise than TLS method.

3.2. Experiment with Real Image Data

In this experiment, the proposed algorithm has been tested with real image data, the "Kitchen" sequence from CMU. Table 1 shows the result of the recovered 3D motion parameters compared to that of TLS. Note that although the ground truth is not known for the real motion, the accuracy of the results can be evaluated by projecting the feature points in the reference image onto the second image plane using the recovered **R** and **t**. The MSE between the true feature points and the corresponding projected ones for each method is calculated and summarized in table 2. Moreover, in order to show the accuracy of the results visually, the actual optical flows (black arrows) and the projected optical flows (white arrows) are depicted and compared in Fig. 5. We see that the recovered optical flows of the proposed algorithm are quite consistent to the actual ones, while many of TLS are not good as indicated in the circles. Thus, from these results, we can conclude that the proposed algorithm is more robust and accurate than TLS method.

4. CONCLUSION

In this paper, we have presented a new 3D camera motion estimation algorithm under a perspective projection. By introducing a weak perspective projection and a coordinate normalization technique, a good initial estimate for the rotational and translational parameters is obtained for the nonlinear system of equations of the true 3D motion, resulting correct recovery of the actual 3D motion parameters. Experimental results demonstrated that the proposed algorithm outperforms the conventional methods even in the presence of severe quantization and measurement noises.

Table 1. The estimation of the Kitchen motion : (a) Proposed algorithm, (b) Total Least Squares(TLS) method.

	Parameters	(a)	(b)
Translation	t_x	-0.062	-0.054
	t_y	0.006	0.005
	t_z	0.998	0.998
Rotation Axis	α	0.014	0.014
	β	0.005	0.004
	γ	-0.015	-0.014

Table 2. Least Mean Square Error.

	Proposed	TLS
Error[pixel]	9.540	21.689

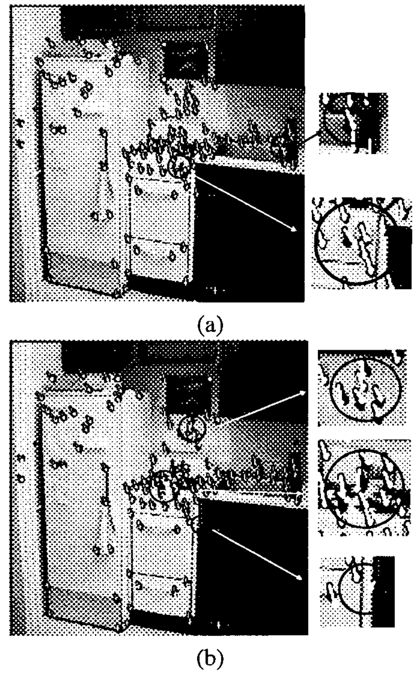

Fig. 5. The comparison of proposed algorithm with TLS for extracted optical flow and estimated optical flow.(a) Proposed algorithm, (b) Total Least Squares method.

5. REFERENCES

[1] J. K. Aggarwal and N. Nandhakumar, "On the Computation of Motion from Sequences of Image - A Review," *Proc. of the IEEE*, vol. 76, no. 8, pp. 917-935, 1988.

[2] Y. Huang et al, "Optic Flow Field Segmentation and Motion Estimation Using a Robust Genetic Partitioning Algorithm," *IEEE Trans. PAMI.*, vol. PAMI-17, no. 12, pp. 1177-1190, 1995.

[3] T. Papadimitriou et al, "Robust Estimation of Rigid-Body 3-D Motion Parameters Based on Point Correspondences," *IEEE Trans. CSVT.*, vol. 10, no. 4, pp. 541-549, 2000.

[4] E. Trucco, and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[5] W. H. Press et al, *Numerical Recipes in C*, Cambridge University Press, 1992.

[6] K. I. Diamantaras and M.G. Strintzis, "Camera Motion Parameter Recovery under Perspective Projection," *in Proc. ICIP.*, pp. 807-810, 1996.

[7] K. I. Dianmantaras et al, " Total Least Squares 3-D Motion Estimation," *in Proc. ICIP.*, pp. 923-927, 1998.