A NEW METHOD FOR CAMERA MOTION PARAMETER ESTIMATION

Yap-Peng Tan *Sanjeev R. Kulkarni* *Peter J. Ramadge*

Department of Electrical Engineering
Princeton University, Princeton, NJ 08544

ABSTRACT

We derive a six parameter system to estimate and compensate the effects of camera motion - zoom, pan, tilt and swing. As compared to other existing methods, this model describes more precisely the effect of different kinds of camera motions. A recursive least-squares estimator has been used to solve for the motion parameters. Experiments suggest that our algorithm converges to satisfactory results when about 10 pairs of corresponding pairs between two image frames are available.

1. INTRODUCTION

Estimating the relative camera motion between two image frames is an important research topic in the areas of computer vision and image coding. In [1], Hoetter models the camera motion as zoom, pan and tilt, while Wu and Kittler [2] model zoom, pan, tilt and swing. Following [1][2], Zakhor and Lari [3] considered a four parameter model to describe the motion effect of zoom, pan, tilt and swing. Unfortunately, the relationships between the model parameters and the physical quantities describing the actual camera motion are not obvious from these models. This makes the estimation of the true physical quantities of camera motion and the analysis of various noise effects difficult. In this paper, we derive a six parameter model to describe the motion of the camera. Besides its computational simplicity and robustness, the proposed model is different from others in that the parameters estimated correspond more precisely to the actual parameters of the camera motion. Under the assumption of small camera motion, our model allows us to estimate the actual pan, tilt, and swing angles, zooming factor, and a scaled focal length of the camera.

2. CAMERA MODEL

Let (X, Y, Z) be a camera coordinate system attached to the camera (Fig 1). The image plane is perpendicular to the Z-axis with its center located at point $(0,0,f)$, where f is the focal length of the camera. Let (x, y) be the corresponding image coordinate system of the image plane. Under perspective projection, a point $P = (X, Y, Z)$ in 3-D space is projected onto the point $p = (x, y)$ on the image, where $x = \frac{fX}{Z}$ and $y = \frac{fY}{Z}$.

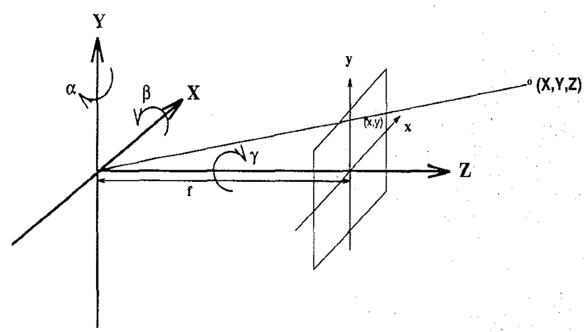

Figure 1: The camera coordinate system, image coordinate system and perspective imaging.

In this section, we model the camera motion between two image frames by the following motion parameters: 1) zoom factor s : ratio of the camera focal lengths between two image frames; 2) pan angle α : rotation angle around the Y-axis; 3) tilt angle β : rotation angle around the X-axis; 4) swing angle γ : rotation angle around the Z-axis; 5) translation vector $\mathbf{t} = (t_x, t_y, t_z)^T$.

Let $\mathbf{P} = (X, Y, Z)^T$ be the camera coordinate of a stationary point $\mathbf{P}$ at time t and $\mathbf{P}' = (X', Y', Z')^T$ be the corresponding camera coordinate at time t'. Between time t and t', the camera undergoes motion described by a change in the above parameters. It is known that any 3-D camera motion can be described by a rotation $\mathbf{R}$ about an arbitrary axis through the origin of the camera coordinate system and followed by a translation $\mathbf{t}$, i.e., $\mathbf{P}' = \mathbf{RP} + \mathbf{t}$. Also, such a 3-D rotation $\mathbf{R}$ can always be represented by successive rotations $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$, $\mathbf{R}_\gamma$ around the Y-,X-, and Z-axes of

Research partially supported by Siemens Corporate Research, Princeton, NJ.

0-8186-7310-9/95 $4.00 © 1995 IEEE

$$
\begin{aligned}
\mathbf{P}' &= \mathbf{R}_\gamma \mathbf{R}_\beta \mathbf{R}_\alpha \mathbf{P} + \mathbf{t} \\
&= \begin{bmatrix} \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & \cos\beta\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ -\cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & cos\beta\cos\gamma & \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\ \sin\alpha\cos\beta & -\sin\beta & \cos\alpha\cos\beta \end{bmatrix} \mathbf{P} + \mathbf{t} \qquad (1)
\end{aligned}
$$

the camera coordinate system respectively, where

$$
\begin{aligned}
\mathbf{R}_\alpha &= \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \\
\mathbf{R}_\beta &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{bmatrix} \\
\mathbf{R}_\gamma &= \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

Combining them together, we have the relationship of equation (1). If the image frames are obtained by sampling at a high enough rate, it is reasonable to assume that the camera undergoes small motions between two image frames. We can then express the image coordinate (x', y') after camera motion in terms of the corresponding image coordinate (x, y) before camera motion and the camera motion parameters

$$\mathbf{z}_i = \mathbf{H}_i \mathbf{a} + \mathbf{n}_i$$

where the vector $\mathbf{a}$ consists of the six intermediate parameters which are functions of the unknown camera motion parameters that need to be estimated; $\mathbf{n}_i$ is a noise vector to account for the neglected higher order terms and measurement error of the corresponding pairs; $\mathbf{z}_i$ is the image coordinate $(x_i', y_i')^T$ after camera motion and $\mathbf{H}_i$ is a 2×6 matrix which entities are functions of image coordinates (x_i, y_i) and (x_i', y_i'). The focal length f (up to a scalar factor), zoom factor s, pan α, tilt β and swing γ angles can be easily retrieved from the six estimated intermediate parameters.

3. LEAST-SQUARES ESTIMATION OF THE CAMERA MOTION PARAMETERS

Having measured m ($m \geq 3$) pairs of corresponding pairs, an estimate for the unknown constant vector $\mathbf{a}$ can be obtained by the least-squares estimator,

$$\hat{\mathbf{a}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{z} \qquad (2)$$

where $\mathbf{H}^T = (\mathbf{H}_1^T \mathbf{H}_2^T ... \mathbf{H}_m^T)$, $\mathbf{z}^T = (\mathbf{z}_1^T \mathbf{z}_2^T ... \mathbf{z}_m^T)$, provided $det(\mathbf{H}^T\mathbf{H}) \neq 0$.

Instead of the previous "batch processing" estimator, we can also use a recursive least-squares estimator,

$$
\begin{aligned}
\hat{\mathbf{a}}_k &= \hat{\mathbf{a}}_{k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{a}}_{k-1}) \\
\mathbf{K}_k &= \mathbf{P}_{k-1}\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \qquad (3) \\
\mathbf{P}_k &= (\mathbf{P}_{k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k)^{-1}
\end{aligned}
$$

where k is the index of the measured corresponding pairs between two images; $\mathbf{R}_k$ is the expected squared error matrix of the neglected higher order terms and measurement error of the corresponding pairs; $\mathbf{K}_k$ is the estimator gain matrix which weights the available innovation $(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{a}}_{k-1})$ and adds to the old estimation $\hat{\mathbf{a}}_{k-1}$; $\mathbf{P}_k$ is the estimation error matrix at the k^{th} measurement.

In our algorithm, we use four corresponding pairs to find an initial estimate of the parameter $\mathbf{a}$ by equation (2). Then, this initial estimate is used to start the recursive least-squares estimator (3). From our experiments, we found that the estimate converges to satisfactory results after about 10 corresponding pairs become available.

4. SIMULATION RESULTS

The algorithm has been tested on both synthetic data, in which we know the camera motion parameters and corresponding pairs exactly, and a variety of real world images with unknown camera motion parameters. For the real world images, we currently pick out some 5x5 pixel blocks with distinct features manually and the corresponding pairs are found by a block matching method searching within windows of certain sizes. Of course, the jobs of feature selection and correspondence establishment can be replaced by the algorithms introduced in [4], [5] or [6].

Figure 2 shows the synthetic data where the camera undergoes a rotation $(\alpha, \beta, \gamma) = (-0.006\pi, -0.005\pi, 0.005\pi)$radians, zooms with a factor $s = 1.15$ and has focal length $f = 200$. Figure 3 is real world images with unknown camera motions. The estimated camera motion parameters are shown in table 1. In order to get some sense of the accuracy of our model and the estimated parameters, non-compensated and compensated frame difference images are also shown in the figures. To reduce edge distortions, bilinear interpolation has been applied for image compensation by using the image intensities of the four nearest neighboring pixels.

5. CONCLUDING REMARKS

From the experimental results, it can be seen that the proposed camera motion model works well and the motion parameters can be efficiently estimated by measuring the corresponding pairs. The fact that the parameters estimated in our method correspond more closely with the actual physical parameters of the camera motion should make the analysis of the sensitivity of this model to various noise effects much easier. Also the computational simplicity and robustness of this method may make it suitable for real time video image stabilization and model-based video coding. If the depth of some scene points is available, we can remove the restriction of small translation motion (but still keep the assumption of small rotation angles), and by similarly approach, we can have a nine parameter camera model. From this model, it seems that we can estimate the zoom factor, pan, tilt, swing angles and the translation vector. We are currently investigating and testing this new model with some image data.

6. REFERENCES

[1] M. Hoetter, "Differential estimation of the global motion parameters zoom and pan," *Signal processing*, Vol. 16, pp.249-265, Mar. 1989.

[2] S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation," *Signal processing: Image communication*, Vol. 2, pp.69-80, 1990.

[3] A. Zakhor and F. Lari, "Edge-based 3-D camera motion estimation with application to video coding," *IEEE Trans. Image processing*, Vol. 2, pp.481-498, Oct. 1993.

[4] J.Q. Fang and T.S. Huang, "A corner finding algorithm for image analysis and registration," *Proc. AAAI-82, Pittsburgh*, Vol. 26, pp.46-49, 1982.

[5] J.Q. Fang and T.S. Huang, "Some experiments on estimation the 3-D motion parameters of a rigid body from two consecutive image frames," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-6, pp.545-554, Sept. 1984.

[6] B. Lukas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. DARPA IU Workshop*, pp.121-130, 1981.

results from figure 2	
motion parameters	estimated motion parameters
$\alpha = -0.006\pi$ (rad)	$\alpha = -0.00589\pi$ (rad)
$\beta = -0.005\pi$ (rad)	$\beta = -0.00509\pi$ (rad)
$\gamma = 0.005\pi$ (rad)	$\gamma = 0.00494\pi$ (rad)
$s = 1.15$	$s = 1.15009$
$f = 200$	$f = 200.089$

results from figure 3	
motion parameters	estimated motion parameters
unknown	$\alpha = 0.00436\pi$ (rad)
	$\beta = -0.00206\pi$ (rad)
	$\gamma = -0.02182\pi$ (rad)
	$s = 1.00495$
	$f = 426.362$

Table 1: Results of estimating motion parameters with proposed algorithm.

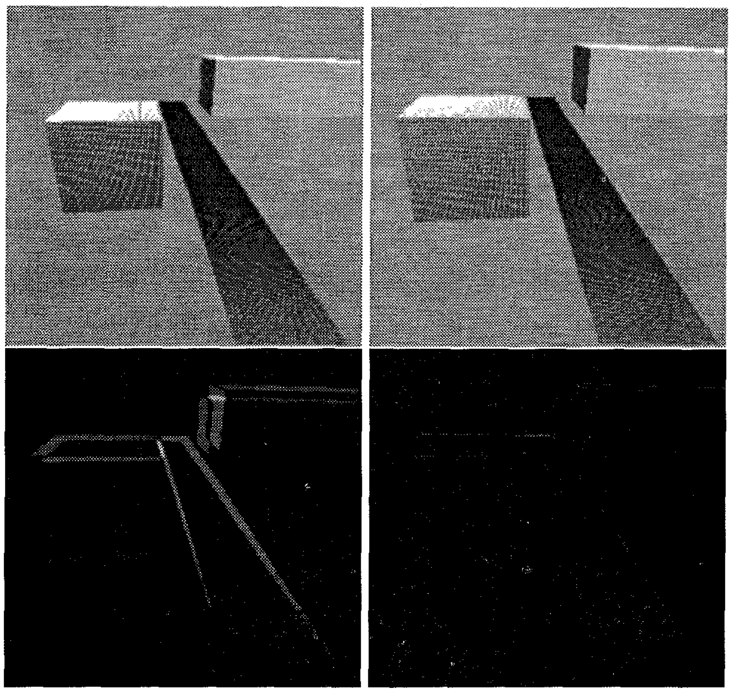

Figure 2: a) top left: first frame image b) top right: second frame image c) bottom left: non-compensated frame difference d)bottom right: compensated frame difference using the proposed algorithm.

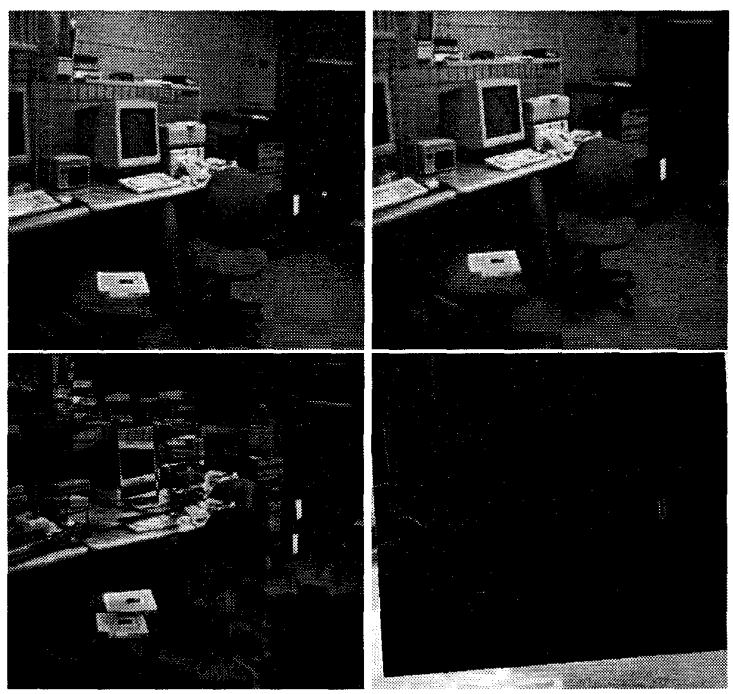

Figure 3: a) top left: first frame image b) top right: second frame image c) bottom left: non-compensated frame difference d)bottom right: compensated frame difference using the proposed algorithm.