PERGAMON

Pattern Recognition 34 (2001) 1573–1583

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Three-dimensional ego-motion estimation from motion fields observed with multiple cameras

Yong-Sheng Chen[a,b], Lin-Gwo Liou[a], Yi-Ping Hung[a,b,*], Chiou-Shann Fuh[b]

[a]*Institute of Information Science, Academia Sinica, 128, Sec 2, Academia Road, Nankang, Taipei 11529, Taiwan*
[b]*Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan*

Received 15 May 2000; accepted 15 May 2000

Abstract

In this paper, we present a robust method to estimate the three-dimensional ego-motion of an observer moving in a static environment. This method combines the optical flow fields observed with multiple cameras to avoid the ambiguity of 3-D motion recovery due to small field of view and small depth variation in the field of view. Two residual functions are proposed to estimate the ego-motion for different situations. In the non-degenerate case, both the direction and the scale of the three-dimensional rotation and translation can be obtained. In the degenerate case, rotation can still be obtained but translation can only be obtained up to a scale factor. Both the number of cameras and the camera placement affect the accuracy of the estimated ego-motion. We compare different camera configurations through simulation. Some results of real-world experiments are also given to demonstrate the benefits of our method.

Keywords: Ego-motion estimation; Multiple sensors; Optical flow

1. Introduction

Motion analysis is concerned with the estimation of the relative motion between an observer and objects. The relative motion is derived from the movement of the observer, the objects, or both. Usually, there are two stages in estimating the motion: first, finding the point correspondences or computing the optical flow field; second, interpreting the motion from the point correspondences [1–6] or the optical flow field [7–11]. Instead of calculating the point correspondences or the optical flow field as an intermediate result, some other methods estimate the motion directly from the spatial and temporal gradients [12,13]. In this paper, we concentrate on the estimation of the so-called ego-motion of an observer moving in a static environment by using the optical flow fields.

Ego-motion provides useful information for human computer interaction and vehicle navigation [14–18]. In the literature, Burger and Bhanu [14] computed the 2-D region of focus of expansion (FOE) as the heading direction of a land vehicle from displacement vectors. Irani et al. [19] removed the effects of rotation by registering 2-D regions. Then they computed the camera translation from the epipolar field. Pei and Liou [18] estimated the vehicle-type motion by using image point and line features. These works estimated the motion according to the camera center, that is, rotation around the axis through the camera center followed by translation. In the application of human computer interaction and navigation, it is more desirable to compute the motion according to the observer's center [15].

One of the major problems in motion recovery is the ambiguity problem. Multiple kinds of motion induce similar optical flow fields and it is difficult to determine the motion from the observed optical flow field. Horn [20] and Brodsky et al. [21] stated that the motion fields and their directions are hardly ever ambiguous, but the ambiguity problem arises if the camera's field of view is

* Corresponding author. Tel.: + 886-2-27883799, ext. 1718; fax: + 886-2-27824814.
E-mail address: hung@iis.sinica.edu.tw (Y.-P. Hung).

0031-3203/01/$20.00
PII: S0031-3203(00)00092-3

small and the variation of the relative depth in the field of view is also small [5,22,23]. In the application of vehicle navigation, for example, consider an airplane with a camera looking down the land or a car with a camera looking far away. If the field of view of the camera is not large enough, the depth map in the view is almost constant. Moreover, it is unsuitable to avoid this problem by using a camera with a large field of view because lens distortion and low resolution may seriously decrease the accuracy of the estimated optical flow.

Another problem in motion recovery is the scaling factor problem concerned with the depth and the translational motion. With only one camera, only the direction of the translation and the relative depth according to the camera center can be estimated [1,8]. The inverse depth and the translation are multiplied together and they can be determined only up to a scale factor.

In this work, we propose a robust method to estimate the three-dimensional ego-motion according to the specified observer center. Several cameras are mounted on the observer and are calibrated [24,25] according to the specified observer center. We use the optical flow fields observed with these cameras to avoid the ambiguity problem. Both the direction and the scale of rotation and translation motion can be obtained by minimizing the proposed residual function of the non-degenerate case. In some special case (degenerate case), for example, when the cameras are not placed well or the observer is undergoing pure translation motion, another residual function can be used to determine the direction and the scale of rotation and the direction of translation.

In the following, we present the proposed method of ego-motion estimation for non-degenerate and degenerate cases in Sections 2.1 and 2.2, respectively. Then we explain why the ambiguity problem can be avoided by using multiple cameras in Section 3. The number of cameras and their placement dramatically affect the accuracy of the estimated motion. We compare the performance of different camera configurations through simulation in Section 4. The results of real-world experiments shown in Section 5 demonstrate the benefits of our method. Finally, conclusions are stated in Section 6.

2. Ego-motion estimation

Consider an arbitrary configuration of K cameras shown in Fig. 1. Without loss of generality, each focal length, f_k, of the kth camera is set to 1. We want to estimate the ego-motion according to the global coordinate system, C_g, attached to the moving observer, where $C_g = \{O, \mathbf{I} = [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3]\}$, O is the origin, $\mathbf{e}_1 = [1,0,0]^{\mathrm{T}}$, $\mathbf{e}_2 = [0,1,0]^{\mathrm{T}}$, and $\mathbf{e}_3 = [0,0,1]^{\mathrm{T}}$. The kth camera coordinate system, C_k, in C_g can be expressed as $C_k = \{\mathbf{b}_k, \mathbf{R}_k = [\mathbf{u}_{k1}|\mathbf{u}_{k2}|\mathbf{u}_{k3}]\}$. The 3×1 vector $\mathbf{b}_k$ denotes the position of O_k and $\mathbf{R}_k$ is a 3×3 orthonormal

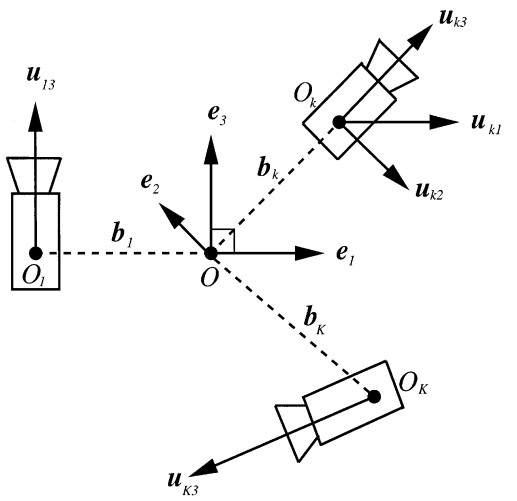

Fig. 1. An arbitrary configuration of K cameras.

matrix. These extrinsic camera parameters $\mathbf{b}_k$ and $\mathbf{R}_k$ of each camera are calibrated beforehand [24,25]. At any time instance, we compute the optical flow fields from the images captured from all the cameras. Let N_k denote the number of image points where the optical flow vectors are calculated in the kth image. Our goal is to compute the 3-D ego-motion according to the global coordinate system from the K optical flow fields.

The 3-D coordinates of a point P in coordinate systems C_g and C_k are $\mathbf{P}$ ($[P_x, P_y, P_z]^{\mathrm{T}}$) and $\mathbf{P}_k$ ($[P_{xk}, P_{yk}, P_{zk}]^{\mathrm{T}}$), respectively. These two coordinate vectors satisfy

$$\mathbf{P} = \mathbf{R}_k\mathbf{P}_k + \mathbf{b}_k. \tag{1}$$

Relative to the global coordinate system C_g, the instantaneous 3-D ego-motion of the point P in the static environment is

$$\dot{\mathbf{P}} = -\omega\times\mathbf{P} - \mathbf{t}, \tag{2}$$

where ω and $\mathbf{t}$ denote the 3-D angular velocity and the translational velocity of the undergoing ego-motion [26]. This 3-D relative motion can also be expressed in the kth camera coordinate system as

$$\dot{\mathbf{P}}_k = -\omega_k\times\mathbf{P}_k - \mathbf{t}_k, \tag{3}$$

where

$$\omega_k = \mathbf{R}_k^{\mathrm{T}}\omega \quad \text{and} \quad \mathbf{t}_k = \mathbf{R}_k^{\mathrm{T}}[(\omega\times\mathbf{b}_k) + \mathbf{t}]. \tag{4}$$

According to the perspective projection camera model, the 3-D point $\mathbf{P}_k$ is projected on the kth image plane at $\mathbf{p}_k$, where

$$\mathbf{p}_k \equiv \begin{bmatrix} p_{xk} \\ p_{yk} \\ 1 \end{bmatrix} = \frac{1}{P_{zk}}\mathbf{P}_k. \tag{5}$$

After temporally differentiating both sides of Eq. (5) and substituting Eq. (3) into $\dot{\mathbf{P}}_k$, we have

$$\mathbf{v}_k \equiv \dot{\mathbf{p}}_k = -(\omega_k \times \mathbf{p}_k) - \frac{\dot{P}_{zk}}{P_{zk}}\mathbf{p}_k - \frac{1}{P_{zk}}\mathbf{t}_k, \tag{6}$$

where $\mathbf{v}_k$ is the optical flow vector at the image point $\mathbf{p}_k$.

Applying cross product ($\times$) by $\mathbf{p}_k$ to Eq. (6), we obtain

$$\mathbf{p}_k \times [\mathbf{v}_k + (\omega_k \times \mathbf{p}_k)] = -\frac{1}{P_{zk}}(\mathbf{p}_k \times \mathbf{t}_k). \tag{7}$$

Then we further apply inner product of $\mathbf{t}_k$ to both sides of Eq. (7) and derive the following fundamental equation which does not contain the unknown depth P_{zk}:

$$\{\mathbf{p}_k \times [\mathbf{v}_k + (\omega_k \times \mathbf{p}_k)]\} \cdot \mathbf{t}_k = 0. \tag{8}$$

The above equation is essentially the infinitesimal version of the epipolar constraint equation [27–29]. Since we want to estimate the 3-D ego-motion ω and $\mathbf{t}$ according to the specified coordinate system, C_g, by using the observations from all the K cameras, the above fundamental equation is re-expressed in terms of ω and $\mathbf{t}$ by using Eq. (4):

$$\mathbf{R}_k\{\mathbf{p}_k \times [\mathbf{v}_k + (\mathbf{R}_k^{\mathrm{T}}\omega \times \mathbf{p}_k)]\} \cdot (\omega \times \mathbf{b}_k + \mathbf{t}) = 0. \tag{9}$$

Once the camera parameters $\mathbf{R}_k$ and $\mathbf{b}_k$ of the kth camera are calibrated and the optical flow vector $\mathbf{v}_k$ at the image position $\mathbf{p}_k$ is estimated, Eq. (9) can be used to determine the 3-D motion parameters, ω and $\mathbf{t}$, without recovering the depth of the image point.

Suppose there are N_k flow vectors associated with the kth camera. We use $\mathbf{p}_{ki}$ and $\mathbf{v}_{ki}$ to represent the ith point and its optical flow associated with the kth camera. Eq. (9) can be rewritten as

$$\mathbf{m}_{ki}^{\mathrm{T}}(\mathbf{h}_k + \mathbf{t}) = 0, \tag{10}$$

where

$$\mathbf{m}_{ki} \equiv \mathbf{R}_k\{\mathbf{p}_{ki} \times [\mathbf{v}_{ki} + (\mathbf{R}_k^{\mathrm{T}}\omega \times \mathbf{p}_{ki})]\}, \tag{11}$$

$$\mathbf{h}_k \equiv \omega \times \mathbf{b}_k. \tag{12}$$

2.1. Non-degenerate case

According to Eq. (10), we define a residual function J_1' which depends on the unknowns ω and $\mathbf{t}$

$$J_1'(\omega, \mathbf{t}) \equiv \sum_{k=1}^{K}\sum_{i=1}^{N_k} \|\mathbf{m}_{ki}^{\mathrm{T}}(\mathbf{h}_k + \mathbf{t})\|^2. \tag{13}$$

Based on the least-squares criterion, the optimal estimates of ω and $\mathbf{t}$ can be obtained by minimizing J_1'. By letting $\partial J_1'/\partial \mathbf{t} = 0$, we have

$$\mathbf{t} = \mathbf{M}^{-1}\mathbf{c}, \tag{14}$$

where

$$\mathbf{M} \equiv \sum_{k=1}^{K}\sum_{i=1}^{N_k} \mathbf{m}_{ki}\mathbf{m}_{ki}^{\mathrm{T}} \quad \text{and} \quad \mathbf{c} \equiv -\sum_{k=1}^{K}\sum_{i=1}^{N_k} \mathbf{m}_{ki}\mathbf{m}_{ki}^{\mathrm{T}}\mathbf{h}_k. \tag{15}$$

In the following, the matrix $\mathbf{M}$ is sometimes written as $\mathbf{M}(\omega)$ to emphasize that $\mathbf{M}$ is a function of ω.

By substituting Eq. (14) into Eq. (13), we have a new residual function J_1 which only depends on the unknown angular velocity ω:

$$J_1(\omega) \equiv -\mathbf{c}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{c} + \sum_{k=1}^{K}\sum_{i=1}^{N_k} \|\mathbf{m}_{ki}^{\mathrm{T}}\mathbf{h}_k\|^2. \tag{16}$$

Therefore, the optimal estimate of ω (denoted by $\hat{\omega}$) based on the least-squares criterion is the one that minimizes the residual function $J_1(\omega)$. Once we have $\hat{\omega}$, the estimate of $\mathbf{t}$, denoted $\hat{\mathbf{t}}$, can be easily obtained by using Eqs. (14) and (15).

2.2. Degenerate case

In some situations, we cannot obtain ω and $\mathbf{t}$ by solving Eqs. (14) and (16).

(1) $K = 1$: Only one camera is used in this case. Eq. (15) becomes $\mathbf{M} = \sum_{i=1}^{N_k}\mathbf{m}_i\mathbf{m}_i^{\mathrm{T}}$ and $\mathbf{c} = -\sum_{i=1}^{N_k}\mathbf{m}_i\mathbf{m}_i^{\mathrm{T}}\mathbf{h} = -\mathbf{M}\mathbf{h}$. Eq. (16) becomes $J_1(\omega) = -\mathbf{h}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{M}\mathbf{h} + \mathbf{h}^{\mathrm{T}}\mathbf{M}\mathbf{h} = 0$. Therefore, we cannot use this residual function, $J_1(\omega)$, when there is only one camera.

(2) $\forall k$, $\mathbf{h}_k = 0$: Eq. (16) becomes $J_1(\omega) = 0$ and useless. Three situations will suffer $\mathbf{h}_k = 0$ for each k. First, $\omega = 0$, that is, there is no rotational motion (pure translational motion). Second, for each k, $\mathbf{b}_k = 0$. This means that O and every O_k coincide at the same point. Third, $\mathbf{b}_k \parallel \omega$ for each k.

(3) $\forall k$, $\mathbf{h}_k = c_k\mathbf{t}$: When $\mathbf{h}_k$ is parallel to $\mathbf{t}$, Eq. (10) becomes $(c_k + 1)\mathbf{m}_{ki}^{\mathrm{T}}\mathbf{t} = 0$. In this case, only the direction of $\mathbf{t}$ can be obtained.

We have to define a new residual function of degenerate case to deal with the above-mentioned situations. Only situation (3) is considered because situation (1) is a special case of situation (2) by letting $O_1 = O$, thus $\mathbf{b}_1 = 0$, and situation (2) is a special case of situation (3) by letting $c_k = 0$. When $\mathbf{h}_k = c_k\mathbf{t}$, Eq. (10) can be reduced into the following form:

$$\mathbf{m}_{ki}^{\mathrm{T}}\mathbf{t} = 0 \quad \text{or} \quad \mathbf{m}_{ki}^{\mathrm{T}}\mathbf{t}_n = 0. \tag{17}$$

The second form of Eq. (17) indicates that only the translational direction is recoverable in these degenerate cases.

Similarly, we can define a residual function J_2' as

$$J_2'(\omega, \mathbf{t}_n) \equiv \sum_{k=1}^{K}\sum_{i=1}^{N_k} \|\mathbf{m}_{ki}^{\mathrm{T}}\mathbf{t}_n\|^2, \tag{18}$$

where $\mathbf{t}_n$ is defined as the unit vector of the direction of translation, $\mathbf{t}$. Expanding Eq. (18), we have

$$J'_2(\omega, \mathbf{t}_n) = \mathbf{t}_n^{\mathrm{T}} \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} \mathbf{m}_{ki} \mathbf{m}_{ki}^{\mathrm{T}} \right) \mathbf{t}_n = \mathbf{t}_n^{\mathrm{T}} \mathbf{M} \mathbf{t}_n, \quad (19)$$

where the Hermitian matrix $\mathbf{M}$ is defined in Eq. (15). When $\mathbf{t}_n \neq 0$, the Rayleigh quotient, $\rho_{\mathbf{M}}(\mathbf{t}_n) = \mathbf{t}_n^{\mathrm{T}} \mathbf{M} \mathbf{t}_n / \mathbf{t}_n^{\mathrm{T}} \mathbf{t}_n = \mathbf{t}_n^{\mathrm{T}} \mathbf{M} \mathbf{t}_n$, is always larger than the smallest eigenvalue, λ, of the Hermitian matrix $\mathbf{M}$. That is, the minimum value of the residual function $J'_2(\omega, \mathbf{t}_n)$ is the smallest eigenvalue of $\mathbf{M}(\omega)$ [29,30].

Given an estimate of ω, the best estimate of $\mathbf{t}_n$ should be the eigenvector of $\mathbf{M}(\omega)$ corresponding to the smallest eigenvalue. We defined a new residual function J_2 which only depends on the unknown ω as

$$J_2(\omega) \equiv \text{the smallest eigenvalue of } \mathbf{M}(\omega). \quad (20)$$

Therefore, the optimal estimate of ω, denoted by $\hat{\omega}$, is the one which minimizes the error function $J_2(\omega)$. The optimal estimate of $\mathbf{t}_n$ (denoted by $\hat{\mathbf{t}}_n$) is the eigenvector of $\mathbf{M}(\hat{\omega})$ corresponding to the smallest eigenvalue.

3. Motion field ambiguity

In this section, we will explain through simulation why the ambiguity problem can be avoided by combining the optical flow fields observed with multiple cameras. Consider a moving vehicle with two cameras mounted on the left and right sides and looking outward as Fig. 2 shows. Two types of motion are under consideration: one is the pure translation motion toward the front direction and the other is the pure rotation motion around the vertical axis of the vehicle. First, let us consider only the left camera (camera 1). The optical flow fields generated by

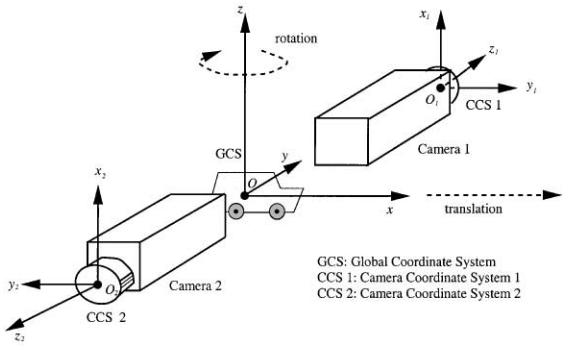

Fig. 2. Two cameras are mounted on the left and right side of the moving vehicle. Two types of motion, pure translation along the X-axis of GCS and pure rotation around the Z-axis of GCS, are under consideration.

the pure translation and the pure rotation are very similar, as shown in Figs. 3(a) and (b), if the field of view of the camera is not large enough (30° in this example) and the depth variation in the field of view is very small. From this single flow field, it is difficult to determine whether the motion is pure translation or pure rotation.

The ambiguity of motion recovery from the optical flow field is illustrated by the following simulation. The vehicle in Fig. 2 is moving straightforward with velocity 10 mm/s. The depth in the field of view is constant (2 m in this simulation). The optical flow field of camera 1 is used to recover the ego-motion of the vehicle by minimizing $J_2(\omega)$ (degenerate case, because pure translation motion is considered). Gaussian noise with three different percentages of the length of the optical flow is applied on the optical flow field. The residual error of function $J_2(\omega)$ is calculated from $\omega_z = -0.5$ to 0.5 ($\omega_x = \omega_y = 0$) and is plotted in Fig. 4. There are two local minima, that is, these two candidate motion are ambiguous. The first one is located near the true motion, $\omega_z = 0$. The residual error at $\omega_z = 0$ increase when larger noise level is applied. The recovered direction of translation (the eigenvector of $\mathbf{M}$) is $[1, 0, 0]^{\mathrm{T}}$ when $\omega_z = 0$ and noise level is 0.

The second local minimum is located near the mistaken motion, $\omega_z = -0.28°/\mathrm{s}$. As the noise level increases, the residual error remains low. The reason is that the noise of flow can be interpreted as the result of the recovered translation vector, $[0, 1, 0]^{\mathrm{T}}$, according to the global coordinate system of the vehicle. To sum up, when the field of view is small and the depth in the field of view is constant, pure translation motion is ambiguous with rotation motion. If the noise of the optical flow is not negligible and we search for the global minimum as the recovered motion, pure translation motion might be interpreted as rotation motion.

Next, let us consider the left and right cameras together on this moving vehicle. If there is only translation, the optical flows observed with the two cameras will be the same in scale but opposite in direction. If there is only rotation, the optical flows will be the same in both the scale and the direction as shown in Fig. 3. Therefore, if we can combine the information contained in the two flow fields appropriately, a more precise and unique motion can be obtained.

The motion fields of the two cameras are used in another simulation and the residual error of $J_2(\omega)$ is plotted in Fig. 5. The only one local minimum (near $\omega_z = 0$) means that there is no ambiguity and the accurate motion can be obtained.

4. Camera placement

Camera placement dramatically affects the robustness and accuracy of the ego-motion estimation with multiple cameras. In Section 2.2, we have described that in some

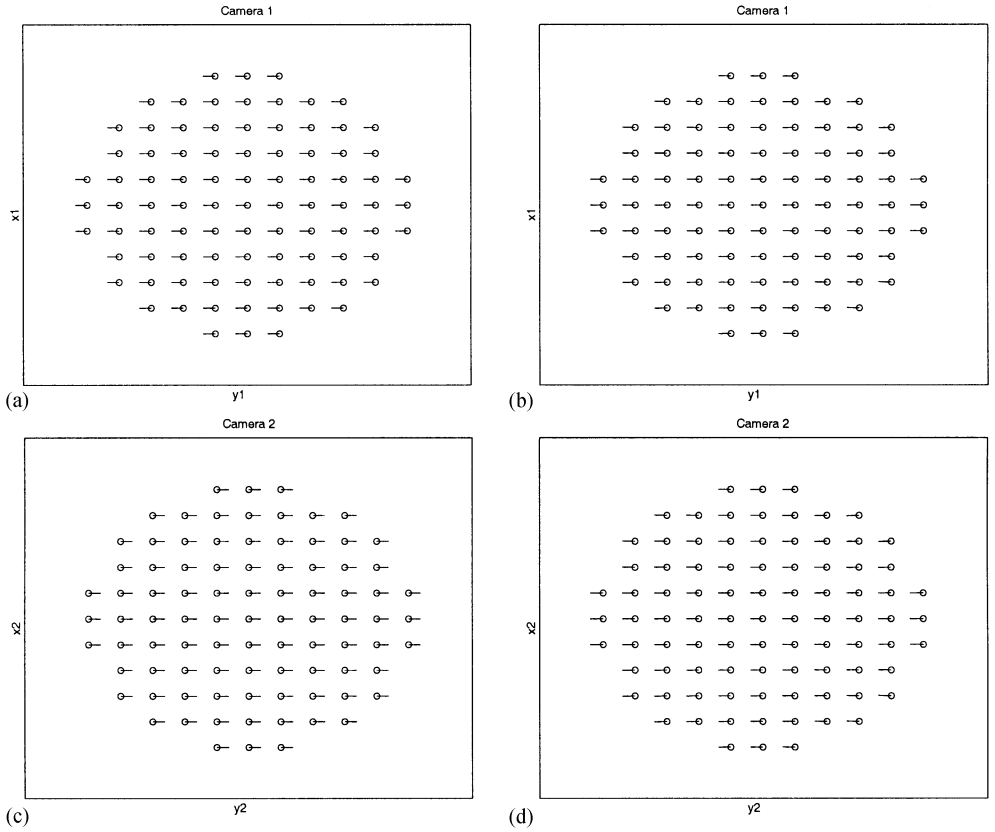

Fig. 3. (a) and (b) are the optical flow fields of the camera 1 when the vehicle is translating (a) and rotating (b). They look very similar and it is difficult to distinguish between them. (c) and (d) are the optical flow fields of the camera 2 when the vehicle is translating (c) and rotating (d). Considering the optical flow fields from both the cameras together, their scales are the same in two kind of motions but the directions are opposite only in pure translation motion.

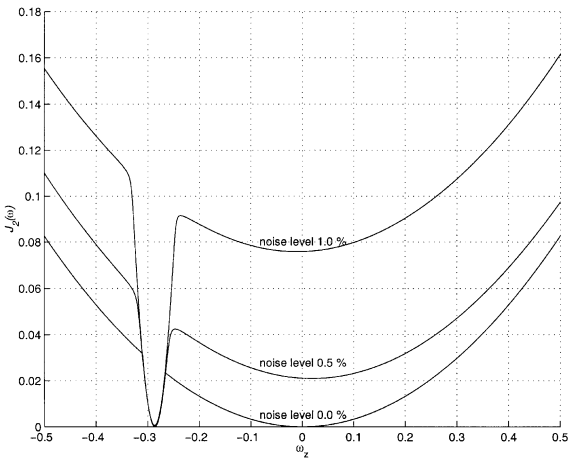

Fig. 4. The residual function $J_2(\omega)$ by using the optical flow field of camera 1 from $\omega_z = -0.5$ to 0.5°/s. These two local minima mean that these two kinds of motion are ambiguous.

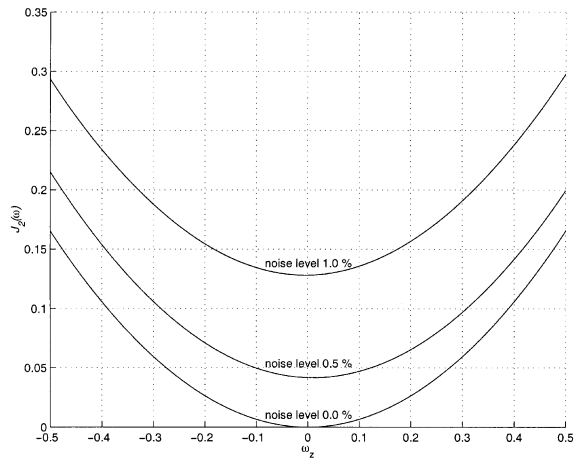

Fig. 5. The residual function $J_2(\omega)$ by using the optical flow fields of cameras 1 and 2 from $\omega_z = -0.5$ to 0.5°/s. There is only one local minimum and the solution is unique.

special configurations of the camera placement, it even turns into degenerate case and only the direction of translation can be estimated. In this section, we will discuss what is the better configuration of the camera placement to obtain more accurate ego-motion.

In Fig. 6, seven cameras are mounted on the observer and the viewing directions of cameras 1 to 7 are Z, $-X$, $-Z$, Z, $-Y$, X, and Y, respectively. The displacements between the origins of the camera coordinate systems and the origin of the observer are $[0,0,100]^T$, $[-100,0,0]^T$, $[0,0,-100]^T$, $[100,0,100]^T$, $[0,-100,0]^T$, $[100,0,0]^T$, and $[0,100,0]^T$, respectively. The optical flow fields of seven kinds of composition of these cameras are used to compute the ego-motion and the accuracy of the result is compared.

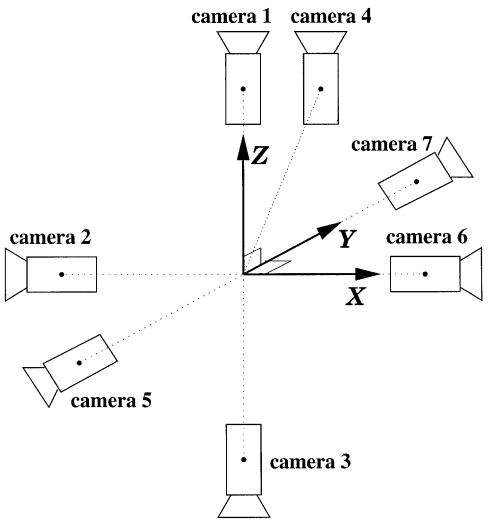

Fig. 6. The seven cameras used for the simulation of camera placement.

In degenerate case, 1000 random trials of 3-D translation motion (the range of each component of the translation motion is from -15 to 15 mm/s) are generated and the optical flow fields for all the cameras are calculated. Gaussian noise of three different noise levels is applied on the optical flow field. Because only the direction of the translation can be estimated in this case, we compare the performance of the seven configurations by calculating the angle between the estimated and the true translation direction, as shown in Table 1. From Table 1, we can observe that: (1) in general, more cameras provide more accurate motion estimation; and (2) orthogonal camera placement (configurations 1 and 5) are better than collinear (configuration 2) and coplanar (configurations 4 and 6) camera placement.

In non-degenerate case, 1000 random trials of 3-D rotation and translation motion are generated. The range of each component of the rotation motion is from -0.5 to 0.5°/s and the range of each component of the translation motion is from -15 to 15 mm/s. Again, the performance of the same seven camera configurations is compared, as shown in Table 2, by computing the angle between the estimated and the true translation direction. Because three-dimensional translation including its scale can be obtained in this case, the distance between the estimated and the true translation motion is also compared in Table 3. Configurations 6 and 7 which use more cameras still can obtain more accurate motion. In this case, collinear (configuration 2) and coplanar (configurations 4 and 6) camera placement can obtain better

Table 1
The average angles (in deg) between the estimated and the true translation directions of seven configurations of camera placement

Configuration	1	2	3	4	5	6	7
Cameras	1,2	1,3	1,4	1,2,3	1,2,5	1,2,3,6	1,2,3,5,6,7
1% noise	0.10	0.37	0.39	0.08	0.06	0.07	0.04
5% noise	0.51	6.56	6.47	0.56	0.32	0.46	0.23
10% noise	1.42	36.04	35.63	1.74	0.67	1.00	0.47

Table 2
The average angles (in deg) between the estimated and the true translation motions of seven configurations of camera placement

Configuration	1	2	3	4	5	6	7
Cameras	1,2	1,3	1,4	1,2,3	1,2,5	1,2,3,6	1,2,3,5,6,7
1% noise	39.91	1.46	18.81	5.41	12.76	0.39	0.17
5% noise	55.71	10.78	46.00	31.95	47.41	5.43	2.71
10% noise	56.15	24.24	52.37	46.55	52.79	13.82	11.83

Table 3
The average distance (in mm) between the estimated and the true translation motions of seven configurations of camera placement

Configuration	1	2	3	4	5	6	7
Cameras	1,2	1,3	1,4	1,2,3	1,2,5	1,2,3,6	1,2,3,5,6,7
1% noise	12.99	5.36	9.74	6.36	8.94	4.44	4.03
5% noise	14.31	12.86	13.72	13.33	13.94	12.79	12.82
10% noise	14.58	14.26	14.30	14.41	14.54	14.27	14.27

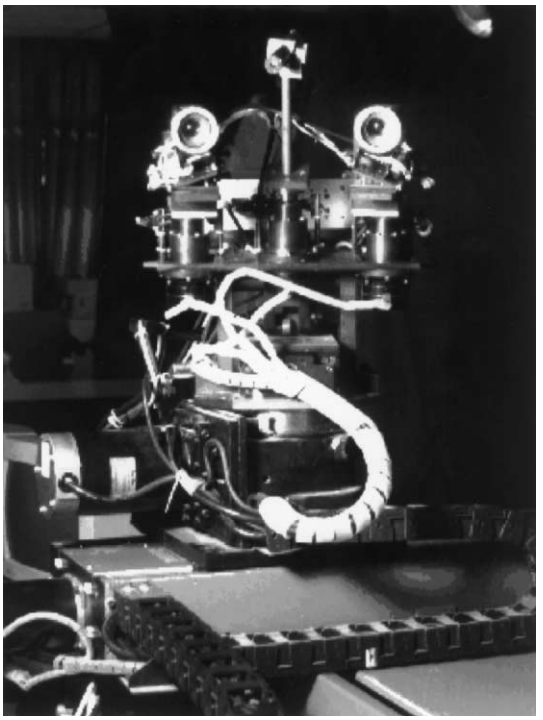

Fig. 7. A picture of the IIS head.

motion estimation than orthogonal ones (configurations 1 and 5).

5. Experimental results

This section shows some results of real experiments. We used a well-calibrated binocular head [25] (referred to as the IIS head) to simulate a moving observer with two cameras mounted on it. The IIS head is built for experiments of active vision, which has four revolute joints and two prismatic joints, as shown in Fig. 7. The two joints on top of the IIS head are for camera verge or gazing. The next two joints below them are for tilting and panning the stereo cameras. All of the above four joints are revolute and are mounted on an X–Y table which is composed of two prismatic joints. The lenses of the binocular head are motorized to focus on objects at different distances.

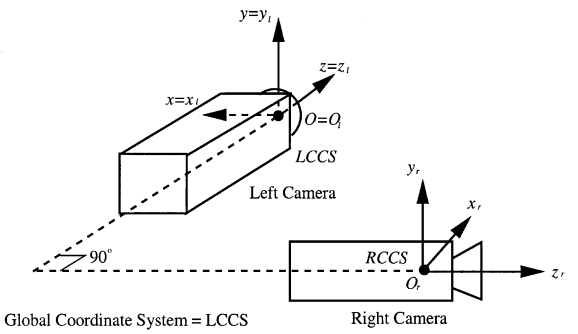

Fig. 8. The coordinate systems and camera configuration of the real experiments.

To simplify the coordinate transform, we let the global coordinate system and the left camera coordinate system be identical. Then the left camera coordinate system ($LCCS$) can be expressed by $LCCS = \{\mathbf{b}_l, \mathbf{R}_l\}$, where $\mathbf{b}_l = \mathbf{0}$ and $\mathbf{R}_l = \mathbf{I}$. We let the angle between the optical axes of left and right cameras be about 90°. Notice that the z-axis of $LCCS$ is the same as the optical axis of the left camera, the x-axis points toward the left side of the left camera, and the y-axis points toward the upper side of the left camera. The focal lengths of both cameras are 25 mm, and the fields of view are 15°. The coordinate systems and the camera configuration are illustrated in Fig. 8.

5.1. Experiment 1

We let the IIS head move forward, such that the left camera of the IIS head looks ahead and the right camera looks to the right. Table 4 lists the true motion parameters used in this experiment. We estimated the ego-motion for three cases: using the left camera only, using the right camera only, and using both the left and right cameras. The scenes viewed from the left and right cameras are shown in Figs. 9(a) and (b), respectively. The optical flow fields observed with the left and right cameras are shown in Figs. 9(c) and (d), respectively. The depth of the scene viewed from left camera is in the range

from 1.3 to 1.5 m, while the depth of the scene viewed from the right camera is about 5 m.

Tables 5 and 6 list the estimates of the rotational parameters and translational parameters. The results show that using both cameras performs better than using only one camera. The performance of using only the left camera is also acceptable because the translation direction is close to the optical axis of the left camera. When

Table 4
True motion parameters used in experiment 1

Rotation			Translation			
ω (deg/frame)			Direction			Mag. (mm)
ω_x	ω_y	ω_z	t_{xn}	t_{yn}	t_{zn}	$\|\|\mathbf{t}\|\|$
0.00	0.00	0.00	− 0.017	0.045	1.00	20.00

Table 5
Rotational parameters estimated in experiment 1

	$\hat{\omega}$ (deg/frame)			Error $\|\|\hat{\omega} - \omega_{true}\|\|$
	$\hat{\omega}_x$	$\hat{\omega}_y$	$\hat{\omega}_z$	
Both	0.017	− 0.034	0.012	0.040
Left	0.00	− 0.052	− 0.052	0.073
Right	− 0.012	0.22	0.00	0.22

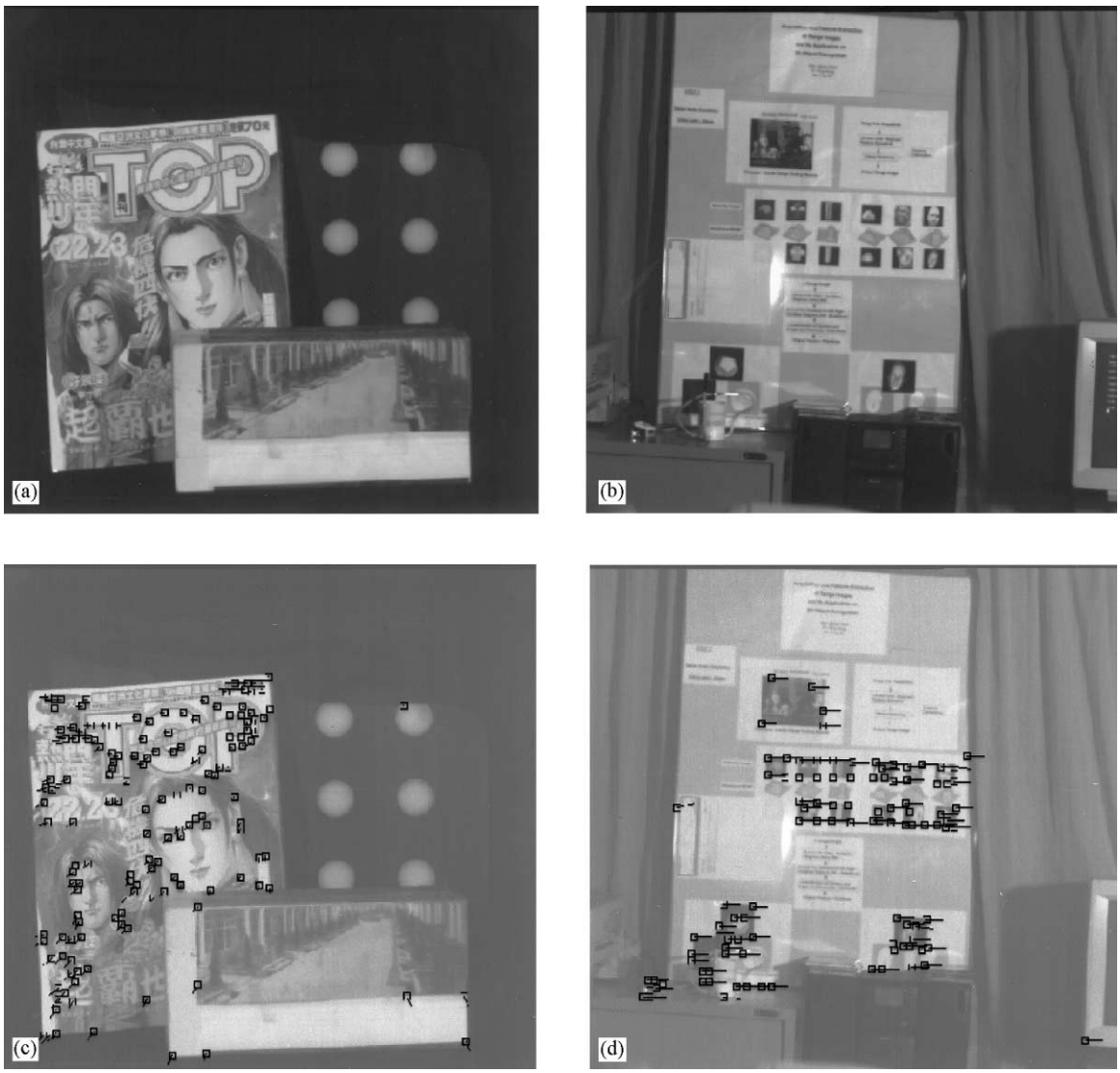

Fig. 9. The images and the optical flow fields used in experiment 1: (a) The scene viewed from the left camera. (b) The scene viewed from the right camera. (c) The optical flow field obtained from the left camera. (d) The optical flow field obtained from the right camera. The optical flow vectors in the figures are enlarged by a factor of two.

Table 6
Translational parameters estimated in experiment 1. $\theta(\hat{\mathbf{t}}_n, \mathbf{t}_{n,true})$ is defined as the angle between $\hat{\mathbf{t}}_n$ and $\mathbf{t}_{n,true}$

	$\mathbf{t}_n$ (translational direction)			Error
	$\hat{t}_{xn}$	$\hat{t}_{yn}$	$\hat{t}_{zn}$	$\theta(\hat{\mathbf{t}}_n, \mathbf{t}_{n,true})$
Both	0.040	0.054	1.00	3.30°
Left	0.062	0.049	1.00	4.54°
Right	− 0.990	− 0.051	0.10	83.39°

Table 7
True motion parameters used in experiment 2

Rotation			Translation			
ω (deg/frame)			Direction			Mag. (mm)
ω_x	ω_y	ω_z	t_{xn}	t_{yn}	t_{zn}	$\|\|\mathbf{t}\|\|$
0.017	0.50	− 0.023	− 0.62	− 0.012	− 0.78	1.89

Table 8
Rotational parameters estimated in experiment 2

	$\hat{\omega}$ (deg/frame)			Error
	$\hat{\omega}_x$	$\hat{\omega}_y$	$\hat{\omega}_z$	$\|\|\hat{\omega} - \omega_{true}\|\|$
Both	0.00	0.52	− 0.029	0.025
Left	0.0057	0.57	0.0057	0.086
Right	0.017	− 0.0057	− 0.011	0.50

Table 9
Translational parameters estimated in experiment 2. $\theta(\hat{\mathbf{t}}_n, \mathbf{t}_{n,true})$ is defined as the angle between $\hat{\mathbf{t}}_n$ and $\mathbf{t}_{n,true}$

	$\mathbf{t}_n$ (translational direction)			Error
	$\hat{t}_{xn}$	$\hat{t}_{yn}$	$\hat{t}_{zn}$	$\theta(\hat{\mathbf{t}}_n, \mathbf{t}_{n,true})$
Both	− 0.052	− 0.034	− 1.00	36.00
Left	0.049	0.055	− 1.00	36.00
Right	0.083	0.99	− 0.095	89.00

only the right camera is used, the ambiguity problem mentioned in Section 3 occurred and the translational motion is mis-classified as rotational motion because the field of view is relatively small and the depth variation in the field of view is also small.

5.2. *Experiment* 2

In experiment 2, we let the IIS head pan with a small angle. Table 7 is the true motion parameters used in this experiment. Again, we estimated the ego-motion by using

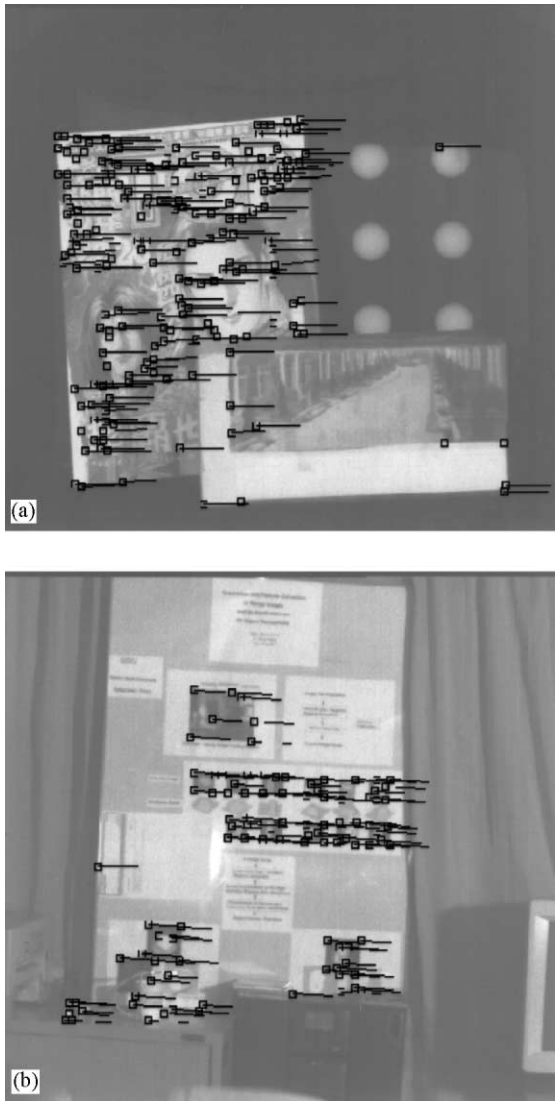

Fig. 10. (a) The optical flow field obtained from the left camera. (b) The optical flow field obtained from the right camera. The optical flow vectors in the figures are enlarged by a factor of two.

the left camera only, the right camera only, and both the left and right cameras, respectively. The scenes viewed from the left and right cameras are the same as Figs. 9(a) and (b). The flow fields observed with left and right cameras are shown in Figs. 10(a) and (b), respectively.

The experimental results of this experiment are given in Tables 8 and 9. As expected, the experiment using both the left and right cameras obtains the most accurate motion. The ambiguity problem occurred again when using the right camera only because the depth of the scene viewed from the right camera was as far as 5 m. Hence, the optical flow field was very similar to the one caused by small pure translation. Notice that the errors

of the translational direction are larger than the ones obtained in Experiment 1. The reason is that the magnitude of translation in this experiment was so small that the estimate of translation was seriously corrupted by the noise of the optical flow field.

6. Conclusions

In this paper, we have proposed a method for 3-D ego-motion estimation using a multiple-camera vision system. This method combines the information contained in the multiple optical flow fields observed with different cameras to avoid the ambiguity problem. Hence, the accuracy of the estimated motion can be improved. Two residual functions are proposed to deal with different cases: non-degenerate case and degenerate case. In the non-degenerate case, 3-D rotation and translation including their scales can be obtained. In the degenerate case, 3-D rotation and the direction of translation can be obtained. Simulations and real experiments show that using multiple cameras can provide more robust and accurate estimate of ego-motion.

One potential application of our multiple-camera approach is the "inside-out" (or "outward looking") head tracker for virtual reality. The current outward looking head tracker requires structured environments, e.g. regular pattern in the ceiling. Our approach does not require specially designed environment, as long as the environment has enough features for computing optical flow.

Acknowledgements

The authors would like to thank the helpful discussion with Dr. Chu-Song Chen and An-Ting Tsao. This work was supported in part by the National Science Council of Taiwan, under Grants NSC 86-2745-E-001-007.

References

[1] R.Y. Tsai, T.S. Huang, Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1) (1984) 13–27.

[2] J. Philip, Estimation of three-dimensional motion of rigid objects from noisy observations, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1) (1991) 61–66.

[3] M.E. Spetsakis, Y. Aloimonos, Optimal visual motion estimation: A note, IEEE Trans. Pattern Anal. Mach. Intell. 14 (9) (1992) 959–964.

[4] J. Weng, P. Cohen, N. Rebibo, Motion and structure estimation from stereo image sequences, IEEE Trans. Robotics Automat. 8 (3) (1992) 362–382.

[5] J. Weng, N. Ahuja, T.S. Huang, Optimal motion and structure estimation, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 864–884.

[6] R.J. Holt, A.N. Netravali, Number of solutions for motion and structure from multiple frame correspondence, Int. J. Comput. Vision 23 (1) (1997) 5–15.

[7] D.T. Lawton, Processing translational motion sequences, Comput. Vision Graphics Image Process. 22 (1) (1983) 116–144.

[8] D.J. Heeger, A.D. Jepson, Subspace methods for recovering rigid motion I: Algorithm and implementation, Int. J. Comput. Vision 7 (2) (1992) 95–117.

[9] R. Hummel, V. Sundareswaran, Motion parameter estimation from global flow field data, IEEE Trans. Pattern Anal. Mach. Intell. 15 (5) (1993) 459–476.

[10] L. Li, J.H. Duncan, 3-D translational motion and structure from binocular image flows, IEEE Trans. Pattern Anal. Mach. Intell. 15 (7) (1993) 657–667.

[11] S. Soatto, P. Perona, Recursive 3-D visual motion estimation using subspace constraints, Int. J. Comput. Vision 22 (3) (1997) 235–259.

[12] S. Negahdaripour, B.K.P. Horn, Direct passive navigation, IEEE Trans. Pattern Anal. Mach. Intell. 9 (1) (1987) 168–176.

[13] B.K.P. Horn, E.J. Weldon Jr., Direct methods for recovering motion, Int. J. Comput. Vision 2 (1) (1988) 51–76.

[14] W. Burger, B. Bhanu, Estimating 3-D ego-motion from perspective image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 12 (11) (1990) 1040–1058.

[15] Y. Liu, T.S. Huang, Vehicle-type motion estimation from multi-frame images, IEEE Trans. Pattern Anal. Mach. Intell. 15 (8) (1993) 802–808.

[16] T. Viéville, E. Clergue, P.E.D.S. Facao, Computation of ego-motion and structure from visual and inertial sensors using the vertical cue, Proceedings of International Conference on Computer Vision, Berlin, Germany, April 1993, pp. 591–598.

[17] A. Giachetti, M. Campani, V. Torre, The use of optical flow for road navigation, IEEE Trans. Robotics Automat. 14 (1) (1998) 34–48.

[18] S.-C. Pei, L.-G. Liou, Vehicle-type motion estimation by the fusion of image point and line features, Pattern Recognition 31 (3) (1998) 333–344.

[19] M. Irani, B. Rousso, S. Peleg, Recovery of ego-motion using region alignment, IEEE Trans. Pattern Anal. Mach. Intell. 19 (3) (1997) 268–272.

[20] B.K.P. Horn, Motion fields are hardly ever ambiguous, Int. J. Comput. Vision 1 (3) (1987) 259–274.

[21] T. Brodsky, C. Fermüller, Y. Aloimonos, Directions of motion fields are hardly ever ambiguous, Int. J. Comput. Vision 26 (1) (1998) 5–24.

[22] G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, IEEE Trans. Pattern Anal. Mach. Intell. 11 (5) (1989) 477–489.

[23] K. Daniilidis, H.-H. Nagel, The coupling of rotation and translation in motion estimation of planar surfaces, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, June 1993, pp. 188–193.

[24] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE J. Robotics Automat. RA-3 (4) (1987) 323–344.

[25] S.-W. Shih, Y.-P. Hung, W.-S. Lin, Calibration of an active binocular head, IEEE Trans. Systems Man Cybernet. 28 (4) (1998) 426–442.
[26] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Vol. 2, Addison-Wesley, Reading, MA, 1993.
[27] H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, Nature 293 (1981) 133–135.
[28] O. Faugeras, Three-Dimensional Computer Vision: a geometric viewpoint, The MIT Press, Cambridge, MA, 1993.
[29] K. Kanatani, Geometric Computation for Machine Vision, Clarendon Press, Oxford, 1993.
[30] B. Noble, J.W. Daniel, Applied Linear Algebra, Prentice-Hall, Englewood Cliffs, NJ, 1988.

About the Author—YONG-SHENG CHEN received his B.S. degree in Computer and Information Science from National Chiao Tung University, Taiwan, in 1993, and M.S. degree in Computer Science and Information Engineering from National Taiwan University, Taiwan, in 1995. He is currently a research assistant at the Institute of Information Science, Academia Sinica, Taiwan, and a Ph.D. student at National Taiwan University. His research interests include computer vision, visual tracking, visual surveillance, and human–machine interaction.

About the Author—LIN-GWO LIOU was born in Taiwan. He received his B.S. degree from the National Chiao Tung University in Taiwan in 1989 and Ph.D. degree from the National Taiwan University in 1995, all in Electrical Engineering. His research interests include motion image analysis, methods for 3-D object reconstruction, pattern recognition in image application.

About the Author—YI-PING HUNG received his B.S. in Electrical Engineering from National Taiwan University in 1982. He received an M.S. from the Division of Engineering, an M.S. from the Division of Applied Mathematics, and a Ph.D. from the Division of Engineering, all at Brown University, in 1987, 1988 and 1990, respectively. He then joined the Institute of Information Science, Academia Sinica, Taiwan, and became a research fellow in 1997. He has been teaching in the Department of Computer Science and Information Engineering at National Taiwan University since 1990, where he is now an adjunct professor. In 1997, he received the Outstanding Young Investigator Award of Academia Sinica. Dr. Hung has published more than 70 technical papers in the fields of computer vision, pattern recognition, image processing, and robotics. In addition to the above topics, his research interests include visual surveillance, virtual reality, human-computer interface, and visual communication.

About the Author—CHIOU-SHANN FUH received the B.S. degree in Computer Science and Information Engineering from National Taiwan University, Taipei, Taiwan, in 1983, the M.S. degree in Computer Science from the Pennsylvania State University, University Park, PA, in 1987, and the Ph.D. degree in Computer Science from Harvard University, Cambridge, MA, in 1992. He was with AT&T Bell Laboratories and engaged in performance monitoring of switching networks from 1992 to 1993. Since 1993, he has been an associate professor in the Computer Science and Information Engineering Department at National Taiwan University, Taipei, Taiwan. His current research interests include digital image processing, computer vision, pattern recognition, and mathematical morphology.