PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 37 (2004) 767–779

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Qualitative estimation of camera motion parameters from the linear composition of optical flow☆

Sang-Cheol Park, Hyoung-Suk Lee, Seong-Whan Lee*

Center for Artificial Vision Research, Department of Computer Science and Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-701, South Korea

Received 9 April 2002; accepted 25 July 2003

Abstract

In this paper, we propose a new method for estimating camera motion parameters based on optical flow models. Camera motion parameters are generated using linear combinations of optical flow models. The proposed method first creates these optical flow models, and then linear decompositions are performed on the input optical flows calculated from adjacent images in the video sequence, which are used to estimate the coefficients of each optical flow model. These coefficients are then applied to the parameters used to create each optical flow model, and the camera motion parameters implied in the adjacent images can be estimated through a linear composition of the weighted parameters.

We demonstrated that the proposed method estimates the camera motion parameters accurately and at a low computational cost as well as robust to noise residing in the video sequence being analyzed.

Keywords: Estimation of camera motion parameters; Video sequences; Optical flows; Linear composition; Linear decomposition

1. Introduction

The estimation of camera pose and motion has become one of the most important aspects of computer vision and object tracking system in recent years. Accurate camera motion is required in the fields of vision-based control, scene analysis, human computer interface, augmented reality (AR), medical imaging and so on. Among these different research topics, AR currently represents one of the greatest challenges in the area of advanced technology.

Video-based AR systems can use image processing and computer vision techniques for the registration of virtual objects. Since they rely heavily on the use of tracking methods, the accurate estimation of camera movements constitutes an essential component of such systems. To this end, we extend the use of camera motion parameters to video-based AR. However, estimating camera motion from video sequences is an arduous process and one that is difficult to accomplish in an efficient manner, since it has the inherent features of being temporal, complex, and massive. Therefore, the ability to reorganize hierarchically and to manage effectively such video sequences is fast becoming a necessity. The goal of this study is to estimate camera motion parameters from video sequences with natural scene features.

A video sequence is generally photographed using a variety of camera motions in order to deliver the information most effectively. A camera has to be oriented according to the movement of the object of interest. The estimation of camera motion parameters is performed in order to obtain, from a particular video sequence, information on the camera motions involved in the video's creation. These camera motions can be classified into seven basic kinds, as shown in Fig. 1. They consist of panning, tilting, zooming and rolling

☆ This research was supported by Creative Research Initiatives of the Ministry of Science and Technology, Korea.

* Corresponding author. Tel.: +82-2-3290-3197; fax: +82-2-926-2168.

E-mail addresses: scpark@image.korea.ac.kr (S.-C. Park), hslee@image.korea.ac.kr (H.-S. Lee), swlee@image.korea.ac.kr (S.-W. Lee).

0031-3203/$30.00
doi:10.1016/j.patcog.2003.07.012

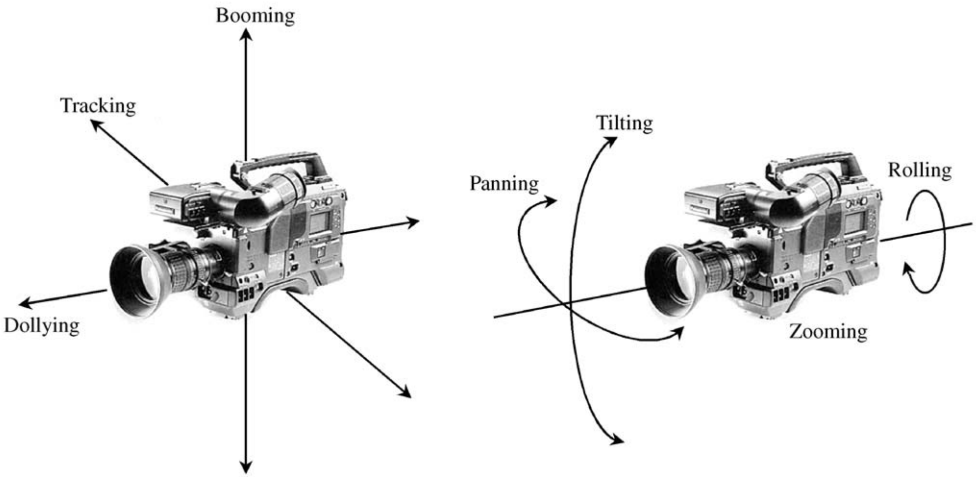

Fig. 1. The basic camera motions.

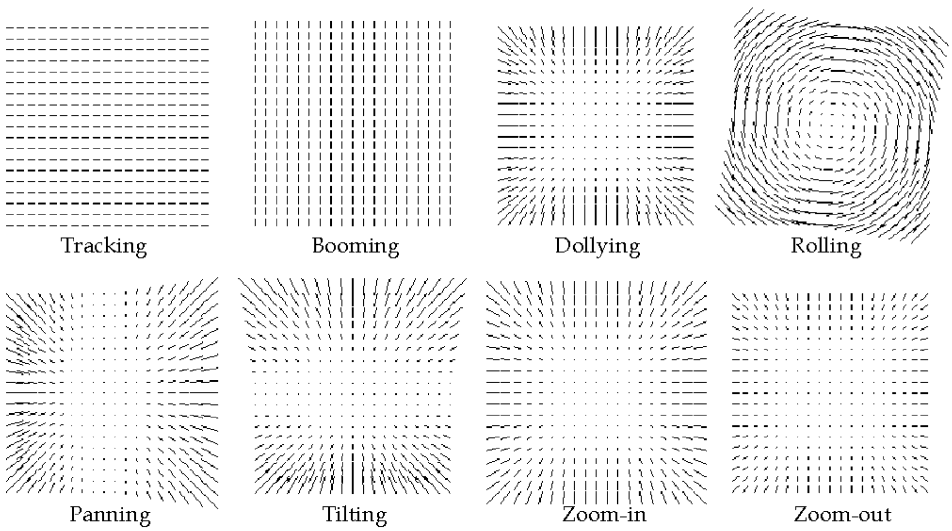

Fig. 2. Optical flows generated by the basic camera motions.

wherein the position of the camera is fixed, and booming, dollying, and tracking, wherein the camera is mobile. Fig. 2 depicts the optical flows generated by these basic camera motions. In this paper, we estimate the camera motion parameters implied in adjacent images through the analysis of the input optical flows using optical flow models.

The optical flows are computed from adjacent images in a video sequence. They are analyzed through the use of optical flow models, which ideally should be established based on noise-free numerical expressions of the optical flows. However, if the amount of noise which they contain is relatively small, the estimated camera motion parameters can still be accurate and even reliable. The proposed whole process is shown in Fig. 3.

This paper is organized as follows. In Section 2, we review related works on estimating camera motion parameters. In Section 3, we provide some essential background information which is required for understanding this paper. In Section 4, we propose a new algorithm which is used to estimate the camera motion parameters. In Section 5, we analyze the results of experiments performed using various synthetic and actual video sequences, in order to evaluate the performance of the proposed algorithm. Finally, our conclusions and some directions for further research are provided in Section 6.

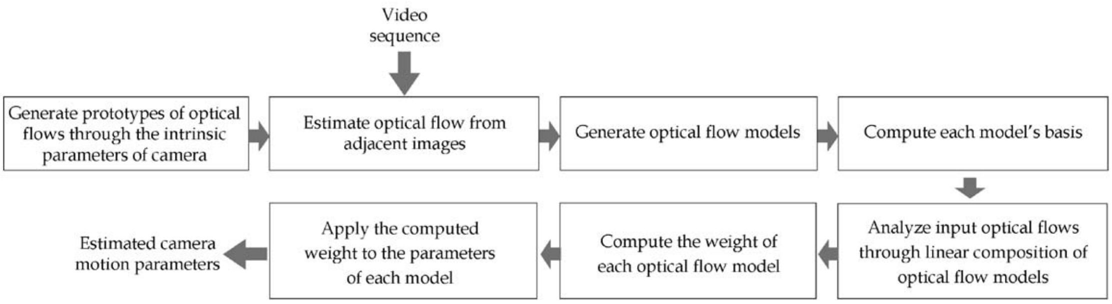

Fig. 3. The proposed process for estimating camera motion parameters.

2. Related works

The estimation of camera motion parameters from a video sequence constitutes one of central themes of computer vision. These parameters are widely used as indices, so as to be able to store and manage large amounts of typical video data in an efficient manner, as well as being employed as references to particular video segments and as a means of tracking objects in a video sequence. Much research is currently being done on the estimation of camera motion parameters, as this technique forms the core of such applications as video compression and augmented reality.

Meada [1] used texture analysis of two-dimensional spatio-temporal images (2DST images) to estimate camera operations in order to describe subscenes in input image sequences. First, a segment is drawn at a position on the screen, and the corresponding segments are then stacked up in time order to form a 2DST image, of which one side lies on the spatial axis and the other on the temporal axis. Each side has a certain length, representing a spatial interval or a temporal period. The movement of contents on the segment within the period can be estimated by investigating the direction of the textures on the 2DST image. The power spectrum is computed by applying a two-dimensional discrete Fourier transform and is used to calculate the direction of the textures in the 2DST images.

Joly et al. [2] generated spatio-temporal images, called simplified X-ray images (SXI), from a video sequence, and analyzed camera motions using information on the angles made by the lines extracted from these SXI. The SXI intensity is quantized in the process of line extraction to reduce the computational cost. The images that include the extracted lines are divided into subunits, and then a Hough transform is performed to obtain the angles made by the lines extracted from these subunits. Utilizing the angles obtained from a particular subunit, they analyzed the camera motions that were employed to generate the video sequence.

Zakhor et al. [3] proposed two models for camera motion analysis. The first model can be applied in situations where there is no camera translation and the motion of the camera can be adequately modeled by the zoom, pan, and rotation parameters. The second model can be applied to situations where the camera is undergoing a translational motion, as well as rotation, zoom and pan. The camera motion parameters are obtained by feeding data corresponding to adjacent images into these two models. They used the binary matching of corresponding edges in correspondence for extracting the camera parameters, instead of the intensity matching technique employed in conventional approaches. The former approach requires one-bit edges to find correspondence between adjacent images while the latter needs one-byte intensity. Therefore, their method has been found to produce a result, which is as good as that obtained from the conventional methods, but at a lower computational cost.

Park et al. [4] estimated the camera motion parameters by fitting correspondence data to a transformation model based on a perspective mapping model and a 3D rotation and zoom operational model. However, 3D translation parameters are not considered since the authors make the assumption that the camera position is fixed. They argued that only those blocks containing a corner point or high texture could convey proper motion information. They first tested whether or not each block carried proper motion information and then estimated the local motion of the blocks that were regarded as conveying proper motion information. Based on the established corresponding data, they used the least-squares estimation technique to compute the camera motion parameters.

Duric et al. [5] drew a 2D histogram based on the optical flows obtained from adjacent frames in a video sequence, observed the relationship between the camera motions and the corresponding changes of the histogram, and analyzed the camera motions implied in the adjacent frames. The rotation of the camera along the optical axis is represented by the rotational changes of the optical flows in the histogram and the translation of various camera motions. If the motion is lateral translation, the histogram is biased in the direction of motion, and if the motion is panning, the histogram is approximately constant.

Akutsu et al. [6] proposed a method based on the analysis of the distribution of motion vectors in Hough space. Each camera motion results in different signature curves in Hough space. The estimation of motion parameters is then based on Hough transformation of the optical flow vectors obtained from the image sequences and the determination of which of the signatures best fits the data in a least-squares sense. However, this technique does not permit the estimation of the magnitude of the motion.

Xiong et al. [7] suggested a method of classifying dominant camera motions in video shots. The method consists of analyzing the optical flow in a decomposed manner. The images are divided into subregions. The projected x and y components of an optical flow are analyzed separately in the different subregions of the images. Different camera motions are recognized by comparing the computed result with the prior known patterns. The proposed method is both efficient and effective because only some mean values and standard deviations are used.

Srinivasan et al. [8] observed that the residual optical flow vectors were parallel when the components of the optical flow, due to camera rotation and zoom, were subtracted. They used an iterative algorithm to minimize the deviations from parallelism of the residual flow vectors. They found r_x, r_y, r_z, and r_{zoom} to be the best estimates of tilt, pan, roll and zoom, respectively. If the magnitudes of the residual flow vectors were significantly compared to those of the original optical flow vectors, the translator component was estimated.

Sudhir et al. [9] classified an optical flow as either singular or nonsingular, according to whether or not the optical flow vanishes at the camera center. A singular flow is subclassified into Z-rotation and Z-translation zoom by using an affine transform. Z-translation and zoom are further discriminated by the mean and variance of the magnitude of $u(x, y)/x$ and $v(x, y)/y$. A nonsingular flow is subclassified into camera translations and rotations by computing the magnitude of the observed optical flow vectors.

3. Background

A camera projects a 3D world point into a 2D image point. The motion of the camera may be limited to a single motion such as rotation, translation, or zoom, or some combination of these three motions. Such camera motion can be represented in the form of optical flows in an image plane.

3.1. Camera mapping

Fig. 4 depicts a general model of a camera with a camera-centered coordinate system. $P = (X, Y, Z)^t$ is a 3D coordinate of the object point P in the 3D camera coordinate system with its center at point O, the focal point, and Z-axis being the optical axis. $p = (x, y)^t$ is a 2D image coordinate of P in the image coordinate system centered at the intersection of the optical axis Z. The image plane is parallel to X- and Y-axis under perspective imaging. The focal length of the camera, f is the distance between the focal point of the camera and the image plane. Eq. (1) represents the relationship between the 3D camera coordinate system and the image coordinate system:

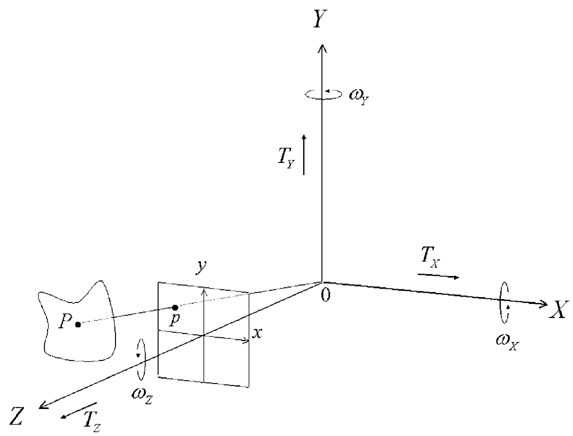

Fig. 4. Camera model.

$$x = f\,\frac{X}{Z}, \qquad y = f\,\frac{Y}{Z}. \tag{1}$$

3.1.1. Camera motion

Camera motion can be considered in two situations: one with the position of the camera being fixed, and the other with the camera being mobile. When the camera position is fixed, the possible camera motions are panning, which is horizontal rotation, tilting, which is vertical rotation, rolling, which is rotation about the optical axis, and zooming, which is modification of the focal distance. When the camera is mobile, the possible movements are tracking, which is a horizontal transverse movement, booming, which is a vertical transverse movement, and dolling, which is a horizontal lateral movement. The camera motion can consist of either a single movement or any combinations of the motions described above. This is defined in Eq. (2) [7,10] below

$$V = T + \Omega \otimes P, \tag{2}$$

where $V=(V_X, V_Y, V_Z)^t$ is the 3D velocity vector at the point where the position vector is $P=(X, Y, Z)^t$, $T=(T_X, T_Y, T_Z)^t$ is a 3D vector that represents the translation of the camera, and $\Omega = (\Omega_X, \Omega_Y, \Omega_Z)^t$ is a 3D vector that represents the angular velocity of the camera rotation.

3.1.2. Optical flow by camera motion

The optical flows generated by the rotation and zoom of the camera are independent of the distance to the object, while those generated by the translation of the camera are dependent on this distance. The optical flows generated by the panning of the camera are composed of horizontally oriented uniform vectors, while those generated by tilting are composed of vertically oriented uniform vectors. The dollying and zooming of the camera produce tangentially and radially oriented vectors, respectively, with the magnitudes of both being proportional to the optical axis.

In the model of the camera centered on its focal point, with its focal length being f, Z represents the optical axis, Y the vertical axis, and X the horizontal axis. T_X, T_Y, and T_Z indicate the translational components of camera velocity along the X-, Y-, and Z-axis, respectively, while Ω_X, Ω_Y, and Ω_Z and the rotational components of the camera's angular velocity along the X-, Y-, and Z-axis, respectively. r_{zoom} represents the camera zoom factor. The optical flows generated at a point on an image plane by the motion of a camera are expressed as vectors with the components, $u(x, y)$ and $v(x, y)$, as shown in Eqs. (3) and (4) [7,11,12] below

$$u(x,y) = -\frac{f}{Z}T_X + \frac{xy}{f}\Omega_X - f\left(1+\frac{x^2}{f^2}\right)\Omega_Y + y\Omega_Z, \quad (3)$$

$$v(x,y) = -\frac{f}{Z}T_Y - \frac{xy}{f}\Omega_Y + f\left(1+\frac{y^2}{f^2}\right)\Omega_X - x\Omega_Z. \quad (4)$$

By including the terms corresponding to camera zoom into the components, u and v, of the optical flows defined in Eqs. (3) and (4), Eqs. (5) and (6) can be obtained

$$u(x,y) = -\frac{f}{Z}T_X + \frac{xy}{f}\Omega_X - f\left(1+\frac{x^2}{f^2}\right)\Omega_Y + y\Omega_Z + f\left[\tan^{-1}\left(\frac{x}{f}\right)\right]\left(1+\frac{x^2}{f^2}\right)r_{zoom}, \quad (5)$$

$$v(x,y) = -\frac{f}{Z}T_Y - \frac{xy}{f}\Omega_Y + f\left(1+\frac{y^2}{f^2}\right)\Omega_X - x\Omega_Z + f\left[\tan^{-1}\left(\frac{y}{f}\right)\right]\left(1+\frac{y^2}{f^2}\right)r_{zoom}. \quad (6)$$

In the above equations, we assumed that the distance between the nearest feature points in the scene is comparable to the change in focal length. We also omitted T_Z, the camera translation along the Z-axis.

3.2. Optical flow on image plane

In this paper, we use optical models to estimate camera motion parameters implied in the adjacent images in a video sequence. The optical models are created by the means of the relational numerical expressions, Eqs. (5) and (6), which describe the optical flows generated on an image plane. When no information is provided on the type of camera used for making the video sequence, it is difficult to generate accurate optical flow models since the focal length included in Eqs. (5) and (6) is unknown. Therefore, we make an assumption that the visual field of the camera is smaller than $20^\circ \times 30^\circ$, in order to derive the relational numerical expressions for the generation of the optical flow models.

In Eqs. (5) and (6), u and v represent the translational and rotational components of camera motion, respectively, as shown in Eqs. (7) and (8) [8] below

$$u = u^{trans} + u^{rot}, \quad (7)$$

$$v = v^{trans} + v^{rot}. \quad (8)$$

Assuming that the visual field of the camera is u^{trans} and v^{trans}, the translational component of camera motion at u and v is sufficiently small, so that Eq. (9) can be derived using orthographic projection.

$$u^{trans} = -\frac{f}{Z}T_X \cong -T_X, \quad (9)$$

$$v^{trans} = -\frac{f}{Z}T_Y \cong -T_Y. \quad (10)$$

With the assumption that the visual field of the camera is u^{rot} and v^{rot}, the rotational component of camera motion at u and v, x/f and y/f can be defined as Eq. (11) shown below [8]

$$\frac{x}{f} \ll 1, \qquad \frac{y}{f} \ll 1. \quad (11)$$

This can be used to derive Eqs. (12) and (13) for u^{rot} and v^{rot}, the rotational components of camera motion

$$u(x,y) = -\frac{f}{Z}T_X + \frac{xy}{f}\Omega_X - f\left(1+\frac{x^2}{f^2}\right)\Omega_Y + y\Omega_Z \cong -f\Omega_Y + y\Omega_Z + xr_{zoom}, \quad (12)$$

$$v(x,y) = -\frac{f}{Z}T_Y - \frac{xy}{f}\Omega_Y + f\left(1+\frac{y^2}{f^2}\right)\Omega_X - y\Omega_Z \cong f\Omega_Z - x\Omega_Z + yr_{zoom}. \quad (13)$$

Although the focal length, f, which is included in Eqs. (12) and (13), cannot be known when no information is provided on the type of camera being used, it only acts as a scaling factor along with the parameter for rotation on the x-axis and the zoom parameter. The focal length can be calculated by estimating the Ω_X and r_{zoom} parameters. Hence, when it is assumed that the visual field of the camera is small, the optical flow created by the panning and tilting of the camera is composed of vectors with uniform magnitude and direction regardless of their position with respect to the image. The optical flow created by zooming, however, is composed of radial vectors whose magnitude is proportional to their distance from the origin.

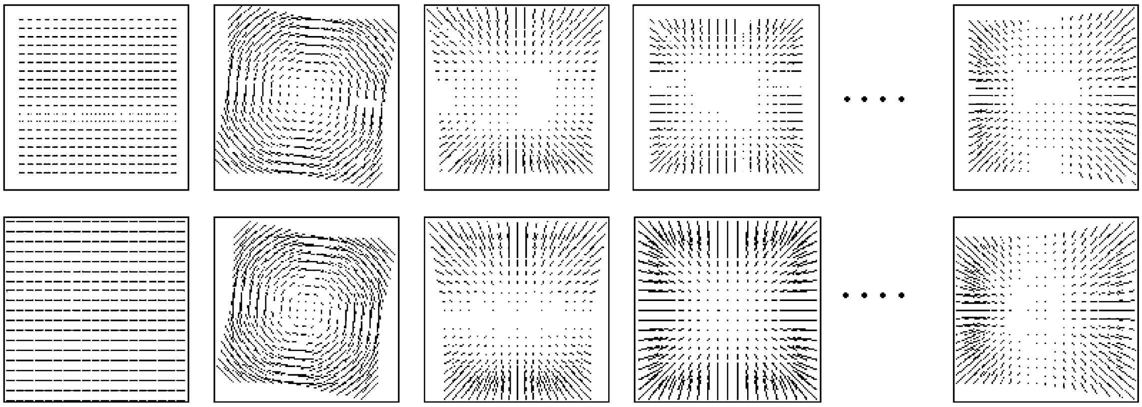

Fig. 5. Prototypes of optical flows generated by the magnitude of each camera movement.

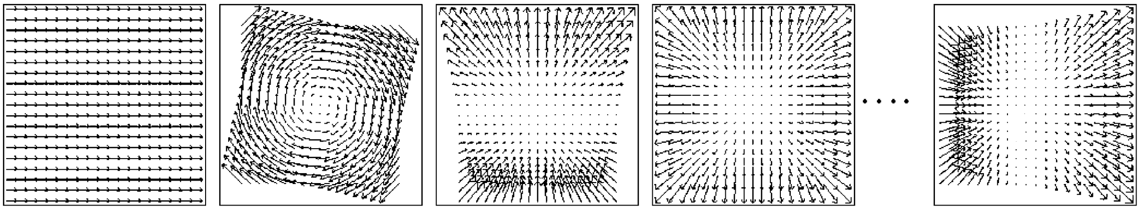

Fig. 6. Prototypes of optical flows generated by the direction of camera movement (arrow direction is positive).

3.3. Prototypes of optical flow models

First, we assume that the intrinsic parameters and the resolution of a specific camera being used are already known. The proposed algorithm for estimating the camera motion parameters establishes prototypes of optical flow models by inputting random values into the translation, rotation, zoom parameters in Eqs. (9)–(13) and the relational numerical expression of the optical flows. These prototypes consist of two different forms, the direction of the optical flows associated with camera movement and the magnitude of the optical flows associated with the velocity at which the camera is moved. For the purpose of analyzing input optical flows, optical flow models have to be generated by obtaining features from the input image and their correspondence to the prototypes of the optical flows whose location is the same as that of the features. Figs. 5 and 6 represent prototypes which consist of optical flow models generated through various camera motions. In these prototypes, the optical flows are created either by a single camera motion, such as translation, rotation, and zoom, or by a combination of these single motions.

4. Algorithm for estimating camera motion parameter

The proposed method for estimating camera motion parameters establishes ideal optical flow models, which are noise-free, by utilizing the numerical expression for the relationship of the optical flows, and creates prototypes of the optical flow models. It then extracts features from the input image, and computes the optical flows for the extracted features. The optical flow models for the analysis of the input optical flows are generated by prototypes of the optical flows whose location is the same as that of the features. The weight of each optical flow model is computed by dividing the input optical flows into the weighted sum of the optical flow models. Finally, the input parameters are computed by calculating the sum of each model's parameters to which its computed weight has been applied. Fig. 7 depicts the algorithm of the proposed method for estimating the camera motion parameters.

4.1. Input optical flow data

In order to estimate the camera motion parameters implied in adjacent images of a video sequence, the optical flows calculated from adjacent images are utilized. For the accurate computation of the optical flows from adjacent images, corner points are identified as features from the input image, based on which the optical flows are computed. The optical flows are then used as input data for the estimation of camera motion parameters implied in adjacent images of the video sequence. Fig. 8 depicts the optical flows calculated at the features identified from the input image. The

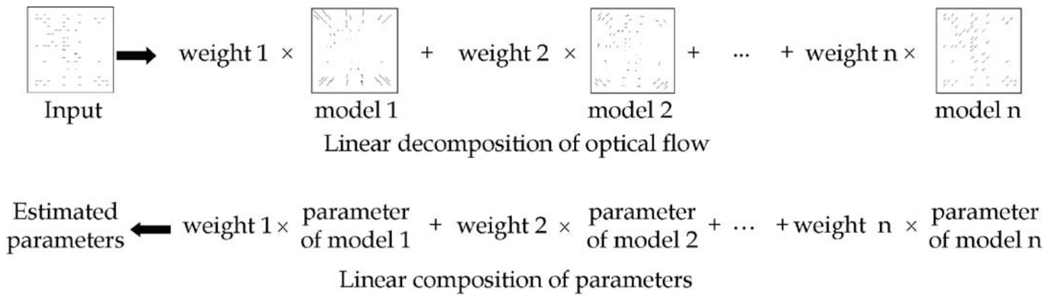

Fig. 7. Estimation algorithm of camera motion parameters.

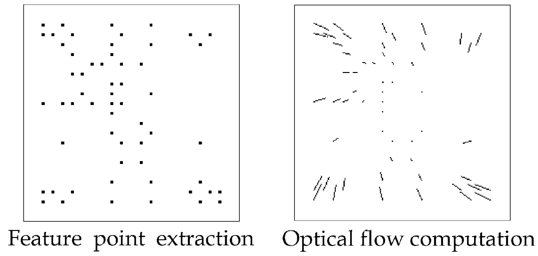

Fig. 8. The computation of optical flows.

location of the features is utilized to generate the optical flow models for the analysis of the input optical flows. In other words, the optical flow models are generated for the analysis of the input optical flows by matching from the database of optical flow models whose location is the same as that of the features.

4.2. *Optical flow models*

Fig. 9 shows the optical flow models generated by the pixelwise correspondence between the input optical flows and the prototypes utilizing the location of the extracted features from the video sequence. These optical flow models express the amount and direction of the camera motion parameters, respectively. So, the input optical flow is represented by the weighted sum of the coefficients of these optical flow models.

4.3. *Linear decomposition of the input optical flows*

The correspondence between the flow field and reference flow field is set in the 2D coordinate system of the x- and y-axis. The flow field parameters correspond to the reference flow field parameters in accordance with their types. The flow field parameters represent the parameters used to generate the optical flow models. The reference flow field parameters are made up of the average of the optical flow models' parameters. The individual parameters composing the set of flow field parameters are encoded into the deformation field with regard to the reference flow field parameters. T is the union set of the optical flow data consisting of the flow field and the parameters used in the generation of the flow field. $\bar{T}$ represents the average of T, $\hat{T} = T - \bar{T}$, the difference of T, and C_T, the covariance of $\hat{T}$. By principal component analysis, a basis transformation is performed to an orthogonal coordinate system formed by eigenvectors t_i of the covariances on data set m [13]:

$$T = \bar{T} + \sum_{i=1}^{m-1} \alpha_i t_i. \tag{14}$$

The probability of the coefficient $\vec{\alpha}$ is defined as

$$p(\vec{\alpha}) \sim \exp\left[-\frac{1}{2}\sum_{i=1}^{m-1}\left(\frac{\alpha_i}{\sigma_i}\right)^2\right]. \tag{15}$$

This section describes a method of finding the coefficient of bases for the estimation of the camera motion parameters implied in a video sequence. We first define an energy function by summing the normalized coefficients, and stipulate that the energy function should be minimized. We then find the coefficient value that minimizes the energy function, using the least-squares minimization. With the optical flows corresponding to the given features, only an approximation of the required parameters can be obtained. Therefore, the goal of this study is to provide an optimal solution in an undetermined condition.

We first define the energy function as the sum of the normalized coefficients, and set the condition that the optical flows corresponding to the features can be fully reconstructed. The energy function $E(\alpha)$ describes the deformation degree of the reference flow field. This paper aims at finding α that minimizes the energy function $E(\alpha)$ for the coefficients satisfying the given condition:

$$\alpha^* = \operatorname*{argmin}_{\alpha} E(\alpha), \tag{16}$$

$$E(\alpha) = \sum_{i=1}^{m-1}\left(\frac{\alpha_i}{\sigma_i}\right)^2, \tag{17}$$

$$\tilde{T}(x_j) = \sum_{i=1}^{m-1} \alpha_i t_i(x_j), \qquad (j = 1, \ldots, n). \tag{18}$$

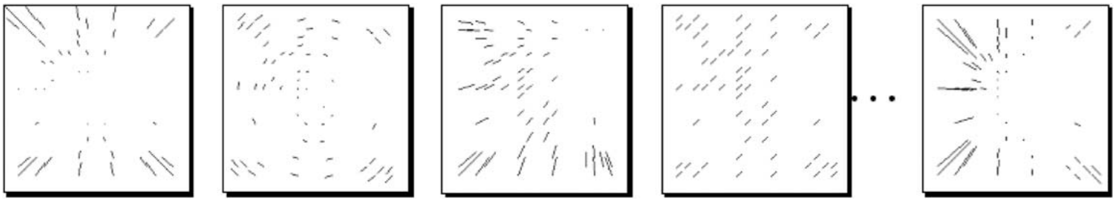

Fig. 9. Optical flow models.

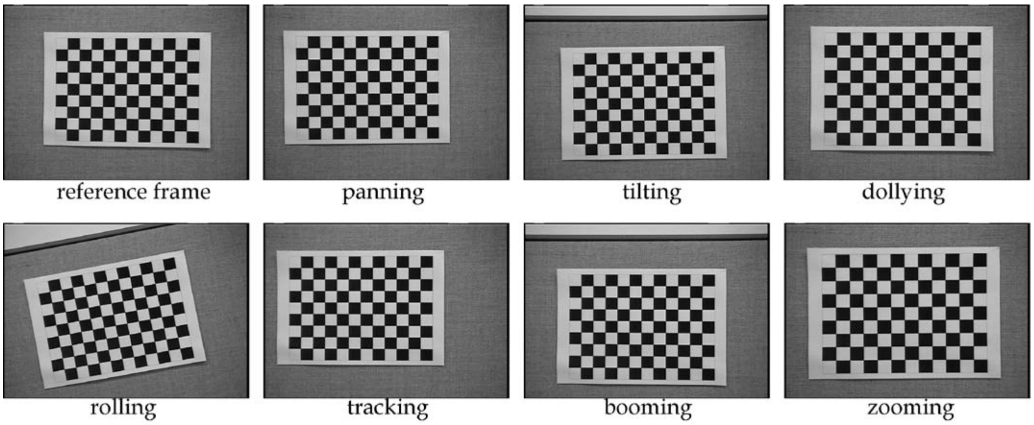

Fig. 10. Synthetic data frames for the accurate camera movement.

Table 1
Experiment on noise-free synthetic data

	T_X	Ω_X	T_Y	Ω_Y	Ω_Z	r_{zoom}
Real	0.2794	0.1974	0.0970	0.4765	−0.9014	0.1421
parameters	0.4768		0.5735		−0.9014	0.1421
Estimated	0.2710	0.2058	0.0902	0.4832	−0.9014	0.1421
parameters	0.4768		0.5734		−0.9014	0.1421
Real	−0.5457	−0.4064	−0.0325	−0.0836	0.1650	0.0184
parameters	−0.9521		−0.1161		0.1650	0.0184
Estimated	−0.5055	−0.4466	−0.0578	−0.0583	0.1650	0.0184
parameters	−0.9521		−0.1161		0.1650	0.0184

T_X: tracking; T_Y: booming; Ω_X: panning; Ω_Y: tilting; Ω_Z: rolling; and r_{zoom}: zooming.

Let $x_1, \ldots, x_n$ be the feature points selected. Since we select only a small number of feature points, n is much smaller than $m-1$.

By using Eqs. (16)–(18), this can be solved though general quadratic programming. To make this problem easier, it is simplified into what can be solved using the least-squares method. Eq. (18) can be equivalently rephrased as follows:

$$\begin{pmatrix} t_1(x_1) & \ldots & t_{m-1}(x_1) \\ \vdots & \ddots & \vdots \\ t_1(x_n) & \ldots & t_{m-1}(x_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{m-1} \end{pmatrix} = \begin{pmatrix} \tilde{T}(x_1) \\ \vdots \\ \tilde{T}n \end{pmatrix}. \tag{19}$$

To exploit the inherent orthogonal nature of the problem, we rewrite Eq. (19) as

$$T\alpha' = \tilde{T}, \tag{20}$$

where

$$T = \begin{pmatrix} \sigma_1 t_1(x_1) & \ldots & \sigma_{m-1} t_{m-1}(x_1) \\ \vdots & \ddots & \vdots \\ \sigma_1 t_1(x_n) & \ldots & \sigma_{m-1} t_{m-1}(x_n) \end{pmatrix},$$

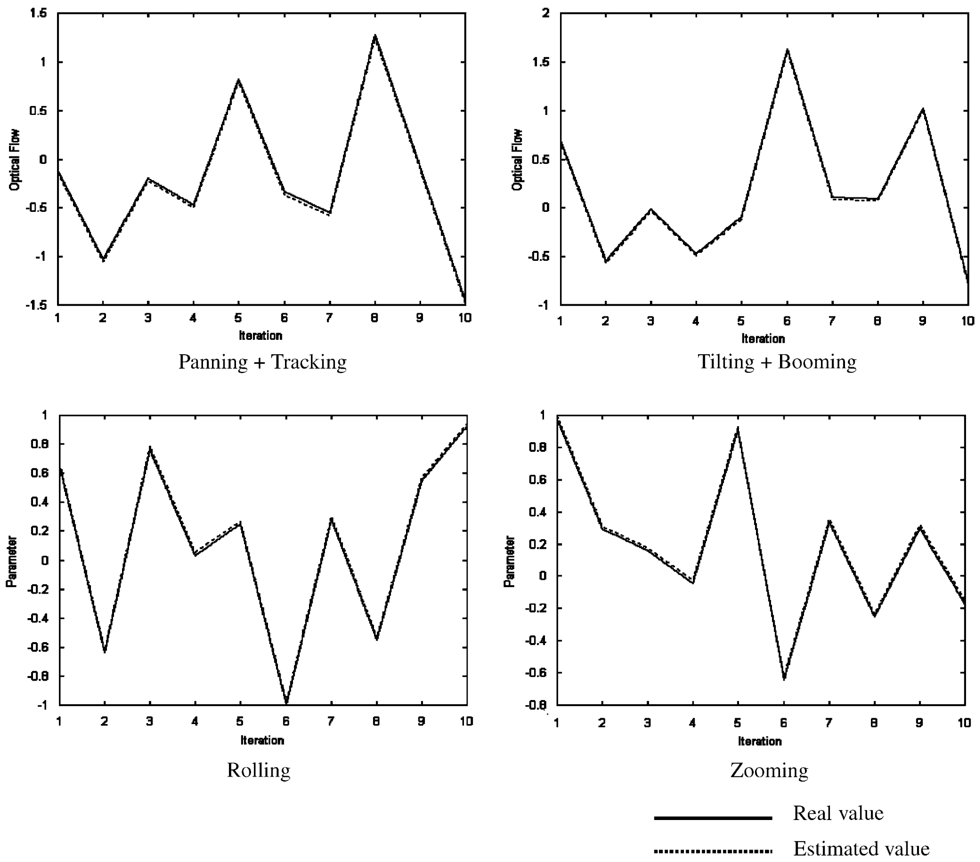

Fig. 11. The difference between the real parameters and the estimated ones.

$$\alpha' = \left(\frac{\alpha_1}{\sigma_1}, \ldots, \frac{\alpha_{m-1}}{\sigma_{m-1}}\right)^{\mathrm{T}}$$

$$\tilde{T} = (\tilde{T}(x_1), \ldots, \tilde{T}(x_n))^{\mathrm{T}}. \tag{21}$$

Let us assume that the row vectors of T are linearly independent. Then, α' can be expressed as

$$\alpha' = T^{+}\tilde{T}. \tag{22}$$

Let T^{+} be the pseudo-inverse of the matrix T. This can be easily obtained through singular value decomposition (SVD). Assuming that the SVD of T is calculated by Eq. (23) and the pseudo-inverse of T can be estimated by Eq. (24) [14]:

$$T = UWV^{\mathrm{T}}, \tag{23}$$

$$T^{+} = VW^{+}U^{\mathrm{T}}. \tag{24}$$

Consequently, the optical flows calculated from adjacent images in the video sequence can be reconstructed by means of Eq. (25) which is derived from Eqs. (14)–(21)

$$T = \bar{T} + \sum_{i=1}^{m-1} \alpha'_i \sigma_i t_i. \tag{25}$$

4.4. Linear composition of parameters

For each model, the weight of its basis, which is calculated through the linear decomposition of the optical flow, as described in the previous section, is applied to the basis of the parameters used for the generation of the model. The weighted bases of the parameters for all the models are summed up and then added to the reference flow field parameters, which eventually provide us with the estimated camera motion parameters for adjacent images in the input video sequence.

For each model, since the matrix T is composed of the optical flows of the model and the parameters which are used for its generation, the basis of the matrix T is that of its optical flows as well as that of the parameters which are used for its generation. Therefore, the estimation of the camera motion parameters for adjacent images can be

Table 2
Experiment on synthetic data with a noise ratio of 7%

	T_X	Ω_X	T_Y	Ω_Y	Ω_Z	r_{zoom}
Real	−0.7283	−0.2026	0.0650	0.4516	−0.2832	−0.4294
parameters	−0.9309		0.5166		−0.2831	−0.4294
Estimated	−0.5197	−0.4004	0.1106	0.4060	−0.2833	−0.4292
parameters	−0.9210		0.5166		−0.2833	−0.4292

Table 3
Experiment on synthetic data with a noise ratio of 8%

	T_X	Ω_X	T_Y	Ω_Y	Ω_Z	r_{zoom}
Real	−0.3409	−0.2042	0.9073	0.0067	−0.7061	−0.4438
parameters	−0.5451		0.9140		−0.7061	−0.4438
Estimated	−0.3078	−0.2722	0.8265	0.1055	−0.7054	−0.4441
parameters	−0.5800		0.9320		−0.7054	−0.4441

Table 4
Experiment on synthetic data with a noise ratio of 9%

	T_X	Ω_X	T_Y	Ω_Y	Ω_Z	r_{zoom}
Real	0.3747	0.2542	0.4291	0.4125	0.7932	0.9960
parameters	0.6289		0.8416		0.7932	0.9960
Estimated	0.2906	0.3159	0.4106	0.4122	0.7903	0.9927
parameters	0.6065		0.8228		0.7903	0.9927

Table 5
Experiment on synthetic data with a noise ratio of 10%

	T_X	Ω_X	T_Y	Ω_Y	Ω_Z	r_{zoom}
Real	0.6125	0.2998	−0.1990	−0.3564	−0.9337	0.9800
parameters	0.9123		−0.5554		−0.9337	0.9880
Estimated	0.4753	0.4114	−0.2927	−0.2347	−0.9289	0.9770
parameters	0.8867		−0.5278		−0.9289	0.9770

defined as

$$P = \bar{P} + \sum_{i=1}^{m-1} \alpha_i' \sigma_i p_i. \tag{26}$$

5. Experimental results and analysis

In order to evaluate the performance of the proposed method for estimating camera motion parameters, experiments were carried out on both synthetic data and actual video data. In the experiment with synthetic data, we use optical flow data generated by various camera motions with the values of the camera motion parameters obtained in a way that substitutes random values for the camera motion parameters in a derived numerical expression.

5.1. Experimental environment

For the experiment with video data, video data containing various camera motions was produced and used. It was taken using a SONY Digital Cam (DCR-PC3), and produced at a rate of more than 15 frames per second using the Optibase MPEG Fusion System MPEG-2 Encoder. Its format is MPEG2 and each frame contains an RGB color image with a resolution of 704×480.

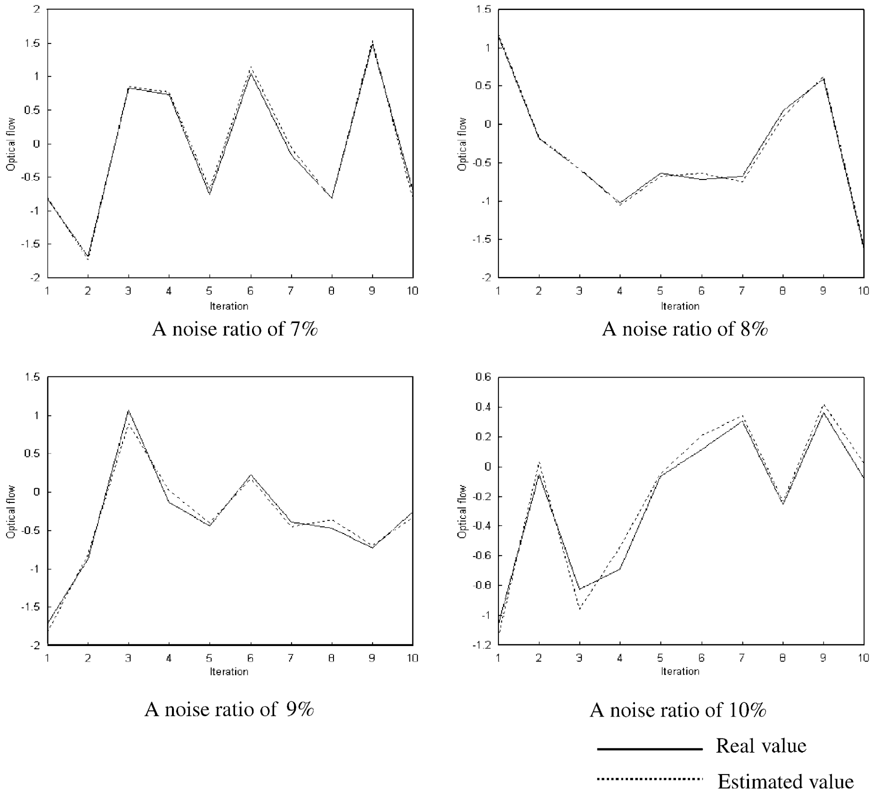

Fig. 12. The performance as to the ratio of noise about a panning parameter.

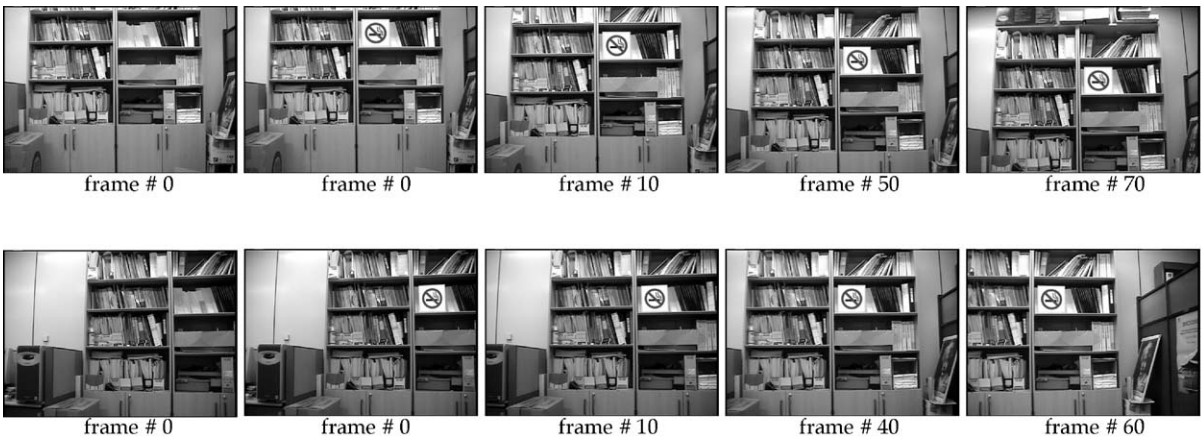

Fig. 13. The insertion of a virtual object using the estimated parameters—tilting camera motion (top) and panning camera motion (down).

5.2. Experimental results

Both synthetic data and real video data were used in our experiments designed to evaluate the proposed method's performance. In the experiment with synthetic data in Fig. 10, the performance of the proposed method was evaluated by observing the difference between the real parameter values and the estimated parameter values. In addition, as a test of noise tolerance, it is done repeatedly with the noise ratio gradually increased. In the experiment with video data, a virtual object is inserted in the video data. The performance of the proposed method was evaluated through the observation of this virtual object moving naturally with the same motions as the background of the video.

5.2.1. Results with noise-free synthetic data

An experiment designed to test the accuracy of the proposed method was made using noise-free synthetic data. The accuracy of the proposed method was evaluated by observing the difference between the real parameter values and the estimated parameter values. Table 1 illustrates the real parameter values and the parameter values estimated by the proposed method.

The camera motions between T_X and Ω_X as well as between T_Y and Ω_Y are very similar. So we can estimate the accuracy of the proposed method using the differences between T_X and Ω_X as well as between T_Y and Ω_Y. As a result of this experiment, the proposed method was shown to be accurate in the estimation of the motion parameters. There was almost no difference between the two parameter values. Also, after we changed the positions and directions of the camera manually according to the camera parameters (panning+tracking, tilting+booming, rolling, zooming), and we estimated the camera parameters in each case. The experimental results are shown in Fig. 11.

5.2.2. Results with noisy synthetic data

An experiment using synthetic data with noise was carried out to test the noise tolerance. This was done repeatedly with the noise ratio gradually increased in the optical flow data used as the input data to the method, in order to determine how robust it was to noise. Tables 2–5 show the performance of the proposed method as to the ratio of noise in the input optical flow data. In Fig. 12, the result of an experiment performed with only one camera parameter being varied is shown in order to analyze its sensitivity to noise.

5.2.3. Results with video sequences

In order to evaluate accurately the camera motion parameters estimated from adjacent images in the video data, a virtual object was inserted into the video data using the estimated parameters. On the basis of this experiment, it was shown that this method provides an accurate means of estimation, through the observation of the virtual object moving naturally with the same motions as the background of the video. Fig. 13 depicts the result of the insertion of a virtual object into the video data using the estimated parameters.

6. Conclusion and future work

In this paper, we proposed a method for estimating camera motion parameters from a video sequence for video-based AR, and demonstrated its performance, in terms of its accuracy in estimating these parameters by means of experiments involving various kinds of video data. With the optical flow data created using the numerical expressions of the optical flows generated by camera motions in an image plane, when this data was noise-free, it was shown that there was almost no difference between the real parameter values and the estimated parameter values. We also proved that even with a large amount of noise in the optical flow data, the estimated parameter values were very close to the real ones. With video sequences, we demonstrated the accuracy of the estimation of the parameters by showing that, after insertion of a virtual object in the video data using the parameters estimated from adjacent image in the video data, the virtual object moved naturally with the same motion as the background of the video.

The proposed method is robust to noise, because it utilizes the ideal model of optical flows that excludes the noise generated as a result of the camera motions. Whether the camera has a single motion or a combination of multiple motions, the proposed method provides an accurate estimation of each motion. Also, the proposed method performs very well even when the camera moves fast.

Further research towards the accurate computation of optical flows from adjacent images will be actively pursued in order to estimate the camera motion parameters even more precisely.

Acknowledgements

This research was supported by a Korea University Grant.

References

[1] J. Maeda, Method for extracting camera operations to describe sub-scenes in video sequences, Proceedings of the SPIE, Vol. 2187, San Jose, CA, 1994, pp. 556–674.

[2] P. Joly. H.-K. Kim, Efficient automatic analysis of camera work and microsegmentation of video using spatiotemporal images, IEEE Signal Process. 8 (4) (1996) 295–307.

[3] A. Zakhor, F. Lari, Edge-based 3D camera motion estimation with application to video coding, IEEE Trans. Image Process. 2 (4) (1993) 481–498.

[4] J. Park, N. Yagi, K. Enami, K. Aizawa, M. Hatori, Estimation of camera parameters from image sequence for model based video coding, IEEE Trans. Circuits Systems Video Technol. 4 (3) (1994) 288–296.

[5] Z. Duric, E. Rivlin, Qualitative description of camera motion from histogram of normal flow, Proceedings of the International Conference on Pattern Recognition, Barocelona, Spain, 2000, pp. 198–201.

[6] A. Akutsu, Y. Tonomura, H. Hashimoto, Y. Ohba, Video indexing using motion vectors, Proceedings of the SPIE, Boston, MA, 1992, pp. 1522–1530.

[7] W. Xiong, J.C.M. Lee, Efficient scene change detection and camera motion annotation for video classification, Comput. Vision Image Understanding 71 (2) (1998) 166–181.

[8] M. Srinivasan, S. Venkatesh, R. Hosie, Qualitative extraction of camera parameters, Pattern Recognition 30 (4) (1997) 593–606.

[9] G. Sudhir, J.C.M. Lee, Video annotation by motion interpretation using optical flow streams, J. Visual Commun. Image Represent. 7(4) (1996) 354–368.
[10] T. Zhang, C. Tomasi, Fast, robust, and consistent camera motion estimation, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, 1999, pp. 164–170.
[11] K.I. Diamantaras, M.G. Strintzis, Camera motion parameter recovery under perspective projection, Proceedings of the International Conference on Image Processing, Lausanne, Switzerland, 1996, pp. 807–810.
[12] E.D. Micheli, V. Torre, S. Uras, The accuracy of the computation of optical flow and of the accuracy of motion parameters, IEEE Trans. Pattern Anal. Mach. Intell. 15 (5) (1993) 434–447.
[13] T. Vetter, T. Poggio, Linear object classes and image synthesis from a single example image, IEEE Trans. Image Process. 19 (7) (1997) 733–742.
[14] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C, Cambridge University Press, Port Chester, NY, 1992.

About the Author—SANG-CHEOL PARK was born in Mokpo, Korea, in 1971. He received his B.S. and M.S. degrees in computer science from Chonnam National University, Kwangju, Korea, in 1996 and 1998. He is a Ph.D. candidate in the Department of Computer Science and Engineering at Korea University, Seoul, Korea. His research interests include computer vision, multimedia, and augmented reality.

About the Author—HYOUNG-SUK LEE was born in Seoul, Korea, in 1975. He received his B.S. degree in computer science from Hansung University, Seoul, Korea, in 1998. He received his M.S. degree in Computer Science and Engineering from Korea University, Seoul, Korea, in 2000. His research interests include pattern recognition and image processing.

About the Author—SEONG-WHAN LEE received his B.S. degree in computer science and statistics from Seoul National University, Seoul, Korea, in 1984, the M.S. and Ph.D. degrees in computer science from KAIST in 1986 and 1989, respectively. From February 1989 to February 1995, he was an assistant professor in the Department of Computer Science at Chungbuk National University, Cheongju, Korea. In March 1995, he joined the faculty of the Department of Computer Science and Engineering at Korea University, Seoul, Korea, as an associate professor, and now he is a full professor. Currently, Dr. Lee is the director of National Creative Research Initiative Center for Artificial Vision Research (CAVR) supported by the Korean Ministry of Science and Technology. He was the winner of the Annual Best Paper Award of the Korea Information Science Society in 1986. He obtained the First Outstanding Young Researcher Award at the Second International Conference on Document Analysis and Recognition in 1993, and the First Distinguished Research Professor Award from Chungbuk National University in 1994. He also obtained the Outstanding Research Award from the Korea Information Science Society in 1996. He has been the co-Editor-in-chief of the International Journal on Document Analysis and Recognition since 1998 and the associate editor of the Pattern Recognition Journal, the International Journal of Pattern Recognition and Artificial Intelligence, and the International Journal of Computer Processing of Oriental Languages since 1997. He was the Program co-chair of the Sixth International Workshop on Frontiers in Handwriting Recognition, the Second International Conference on Multimodal Interface, the 17th International Conference on Computer Processing of Oriental Languages, the Fifth International Conference on Document Analysis and Recognition, and the Seventh International Conference on Neural Information Processing. He was the conference co-chair of the Third International Workshop on Document Analysis Systems, the First IEEE International Workshop on Biologically Motivated Computer Vision, the First International Workshop on Pattern Recognition with Support Vector Machines, the Second International Workshop on Biologically Motivated Computer Vision, and the 6th IEEE International Conference on Automatic Face and Gesture Recognition. He served on the program committees of several well-known international conferences. He is a fellow of IAPR, a senior member of the IEEE Computer Society and a life member of the Korea Information Science Society and the Oriental Languages Computer Society. His research interests include pattern recognition, computer vision and neural networks. He has more than 200 publications on these areas in international journals and conference proceedings, and authored 10 books.