PERGAMON

Pattern Recognition 34 (2001) 1701–1712

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Multi-image matching for a general motion stereo camera model

Ja Seong Ku[a], Kyoung Mu Lee[b], Sang Uk Lee[a,*]

[a]School of Electrical Engineering, Seoul National University, Kwanak P.O. Box 34, Seoul, 151-742, South Korea
[b]Department of Electronics and Electrical Engineering, Hong-Ik University, 72-1 Sangsu-dong, Mapo-gu, Seoul, 121-791, South Korea

Received 19 March 1999; received in revised form 26 January 2000; accepted 8 June 2000

Abstract

This paper presents an accurate and robust motion stereo algorithm employing multiple images, taken under a general motion. The object functions for individual stereo pairs are represented, in terms of the depth, then these object functions are integrated, considering the positions of cameras and the shape of the object functions. By integrating the general motion stereo images, we not only reduce the ambiguities in correspondence, but also improve the accuracy of reconstruction. Also, by introducing a weight for each object function and selecting only reliable object functions, we can improve the accuracy greatly and eliminate false matches, caused by occlusion or noise. Experimental results on the synthetic and real data sets are presented to demonstrate the performance of the proposed algorithm.

Keywords: Stereo; Lateral motion stereo; Axial motion stereo; Multi-image matching; Multiple-baseline stereo

1. Introduction

The problem of extracting the 3-D information from images has been one of the most challenging and important problems in computer vision. Among the existing 3-D reconstruction techniques, shape from stereo, which extracts the 3-D information from multiple images using geometric constraints between corresponding features, has been investigated quite intensively by many researchers [1,2]. Especially, the techniques using stereo images taken by a single moving camera, known as the motion stereo, have received much attention, due to its practicability and accuracy. These techniques can be readily applied to target tracking, navigation, 3-D recognition and so on.

Among many motion stereo techniques, two typical techniques, namely lateral and axial motion stereo, have been intensively investigated [3,4]. The lateral motion stereo acquires images, by moving a camera in the direction perpendicular to its optical axis, while the axial motion stereo moves in the parallel direction. It is known that the axial motion stereo has several advantages over the lateral motion stereo, such as small search range for finding the corresponding features and the low probability of missing them [3]. However, it provides relatively poor estimate, compared with the lateral motion stereo, mainly because of the large error over the region near the center of an image. Although the lateral and axial motion stereo methods are relatively easy to analyze, those methods are quite restrictive in practical applications, since it is not easy to maintain the camera motion perpendicular or parallel to the optical axis in many real situations.

There also have been many stereo techniques using multiple images [5–11]. Most of these techniques find the corresponding points in each stereo image pair and then select the correct combinations for the corresponding points. These techniques are simply the extension of the conventional binocular stereo technique. In contrast, there have been techniques which use multiple images

* Corresponding author. Tel.: +82-2-880-8401; fax: +82-2-880-8201.
E-mail address: sanguk@sting.snu.ac.kr (S.U. Lee).

0031-3203/01/$20.00
PII: S0031-3203(00)00110-2

simultaneously to find the corresponding points. Tsai [5] proposed an algorithm using multiple images to find the sharper extrema of an object function for the corresponding points. Tsai positioned eight cameras at the specific points, called the potential conjugate points on multiple perspective views. Okutomi et al. [6] and Kanade [7] presented a method that uses multiple stereo pairs with various baselines, obtained by lateral displacement of a camera. In these methods, they represented the sum of squared difference (SSD) for each stereo pair, with respect to the inverse distance, rather than the disparity. Then, the sum of the resulting SSD functions is used for an object function to find the correct depth values, demonstrating that [6,7] could remove the ambiguity and improve the accuracy.

However, notice that these methods also assume the restrictive conditions on the camera position, i.e., the motion should be lateral. This problem could be solved by using the operation of rectification [2], but it is still difficult to combine the various kinds of images. Therefore, this paper presents an accurate and robust motion stereo algorithm based on multiple images, taken under a general motion, including both the rotational and translational motion. By using multiple stereo images taken under a general motion in batch mode, we not only reduce the ambiguities in finding the correspondence, but also improve the accuracy of the reconstruction significantly.

The object functions for individual stereo pairs are represented, in terms of the depth, then these object functions are combined, considering the positions of cameras and the shape of the object functions. Then, by introducing a weight for each object function, according to its measured reliability, and selectively employing nonoccluded object functions, we can improve the accuracy of the depth estimate significantly and eliminate the false matches, caused by occlusion or noise. But, in this paper, it is assumed that the intrinsic and extrinsic parameters of camera are already available.

2. Multi-image matching for a general motion stereo

Similar to the method proposed by Okutomi et al. [6], we first represent the SSD, the object function, in terms of the depth, so that the results of each stereo image pair can be integrated in a single framework. Since we consider the general motion of a camera, it may include the axial, lateral, and rotational motions simultaneously. In this case, some feature points in a reference image may no longer exist in other images, due to occlusion or noise. Also, a specific camera configuration could show that the obtained image contains no reliable depth information for certain feature points. In this section, thus, in order to solve these problems and enhance the performance of the motion stereo, we analyze the reliability of the SSD function for a pair of images, according to the geometrical configurations of the camera, and propose a new object function employing only confident SSDs, which are appropriately weighted, according to the degree of their reliability.

2.1. Models and object functions

In this paper, we employ an area-based matching technique for the correspondence problem, and the SSD is used for the correlation measure between the two corresponding windows. When the SSD is calculated under a general motion stereo, the size and shape of corresponding windows in the reference and the other image could be different, due to dilation, rotation and perspective distortion, caused by the general motion of camera. Consequently, a rectangular window in reference image appeared as a general quadrangle in another image. Thus, we employ a variable window method in which the size and shape of the corresponding window varies, according to the relative position of the camera. Moreover, a sub-pixel registration technique is used to improve the precision.

Let us consider a projection of a 3-D space point onto an image plane through a camera. The projection is dependent on both internal and external factors. Typical external factors are the translation and rotation of the camera, with respect to the world coordinate, while internal factors include the focal length, zoom, the image offset, etc. Ignoring additional optical distortions, such as the lens aberrations, a practical pin-hole camera model can be represented by using a homogeneous coordinate system, in which the space point $\mathbf{X} = (X, Y, Z)^{\mathrm{T}}$ and the projected image point $\mathbf{x} = (x, y)^{\mathrm{T}}$ are related by the following linear equation:

$$S\mathbf{m} = \mathbf{P}[\mathbf{R}, \mathbf{t}]\mathbf{M} = \mathbf{P}[\mathbf{R}\mathbf{X} + \mathbf{t}], \tag{1}$$

where $\mathbf{m} = [\mathbf{x}^{\mathrm{T}}, 1]^{\mathrm{T}}$, $\mathbf{M} = [\mathbf{X}^{\mathrm{T}}, 1]^{\mathrm{T}}$,

$$\mathbf{P} = \begin{bmatrix} f\alpha_u & 0 & u_0 \\ 0 & f\alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{r}_1^{\mathrm{T}} \\ \mathbf{r}_2^{\mathrm{T}} \\ \mathbf{r}_3^{\mathrm{T}} \end{bmatrix}, \quad \text{and} \quad \mathbf{t} = [t_x, t_y, t_z]^{\mathrm{T}},$$

where α_u and α_v are the scaling factors, u_0 and v_0 are the offset parameters, $\mathbf{r}_i = [r_{i1}, r_{i2}, r_{i3}]^{\mathrm{T}}$, $i = 1, 2, 3$ are the rotation parameters, and t_x, t_y, and t_z are the translation parameters, respectively. Note that the scaling factor S is the distance (depth) from the image plane to the object point.

Now, consider a camera under a general motion, including both the translational and rotational motion simultaneously, as shown in Fig. 1. Suppose that $N + 1$ images are taken at the optical center positions of the

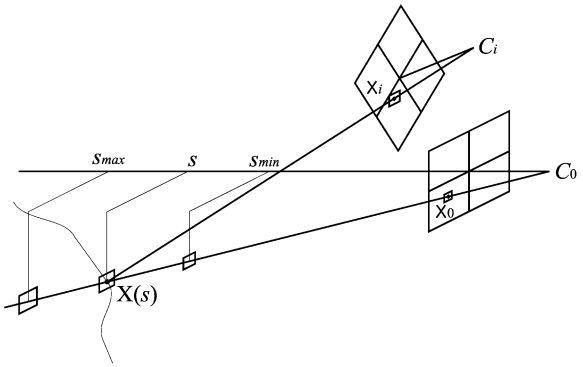

Fig. 1. General motion stereo camera model.

camera, $\mathbf{C}_i$, $i = 0, \ldots, N$, where $\mathbf{C}_0$ denotes the reference image. It is assumed that the rotation matrix, and the translation vector, denoted by $\mathbf{R}_i$ and $\mathbf{t}_i$ for each ith camera position, with respect to the world coordinates, are known. Note that since the projection matrix $\mathbf{P}$ is composed of the internal parameters only, it is constant for all the projected images.

Let $f_0(\mathbf{x}_0)$ and $f_i(\mathbf{x}_i)$ be the intensity functions of the reference and the ith target image, respectively. Then, the SSD function between the reference image and the ith image for a point $\mathbf{x}_0$ in the reference image can be defined, in terms of the depth S for $\mathbf{x}_0$, as

$$SSD_i(\mathbf{x}_0, S) = \sum_{\mathbf{r}_0 \in W} \{f_0(\mathbf{r}_0) - f_i(\mathbf{r}_i(S))\}^2, \tag{2}$$

where W and $\mathbf{r}_0$ denote a local window surrounding the point $\mathbf{x}_0$ in the reference image and a point in the window, respectively, and $\mathbf{r}_i(S)$ denotes the corresponding point to $\mathbf{r}_0$ in the ith image.

Note that for a given point $\mathbf{r}_0$ in the reference image, the corresponding object coordinates can be determined from Eq. (1):

$$\mathbf{X}(S) = \mathbf{R}_0^{-1}[S\mathbf{P}^{-1}\mathbf{m}_0 - \mathbf{t}_0], \tag{3}$$

where $\mathbf{m}_0 = [\mathbf{r}_0^{\mathrm{T}}, 1]^{\mathrm{T}}$.

Then, the corresponding point $\mathbf{r}_i(S)$ projected onto the ith image can be obtained from

$$\mathbf{r}_i(S) = \left[\frac{U_i(S)}{S_i(S)}, \frac{V_i(S)}{S_i(S)}\right]^{\mathrm{T}}, \quad i = 1, \ldots, N, \tag{4}$$

where

$$[U_i(S), V_i(S), S_i(S)]^{\mathrm{T}} = \mathbf{P}[\mathbf{R}_i\mathbf{X}(S) + \mathbf{t}_i] = \mathbf{P}[\mathbf{R}_i\mathbf{R}_0^{-1}(S\mathbf{P}^{-1}\mathbf{m}_0 - \mathbf{t}_0) + \mathbf{t}_i]. \tag{5}$$

It is noted that the corresponding image point $\mathbf{r}_i(S)$ on the ith image is dependent on S, that is, the depth of the reference point $\mathbf{x}_0$.

Now, by summing all the individual SSDs between the reference image $\mathbf{C}_0$ and $\mathbf{C}_i$, $i = 1, \ldots, N$, we define the sum of SSD (SSSD) for the point $\mathbf{x}_0$, in terms of S, as

$$SSSD(\mathbf{x}_0, S) = \sum_{i=1}^{N} SSD_i(\mathbf{x}_0, S) = \sum_{i=1}^{N} \sum_{\mathbf{r}_0 \in W} \{f_0(\mathbf{r}_0) - f_i(\mathbf{r}_i(S))\}^2. \tag{6}$$

Thus, by finding the value S which minimizes Eq. (6), we can reconstruct the depth information for the point $\mathbf{x}_0$, with respect to the reference image plane $\mathbf{C}_0$. In practice, if the approximate range of the object position is known, it can be easily determined by varying S incrementally by ΔS within the search range $[S_{min}, S_{max}]$.

2.2. Elimination of ambiguity using the weighted sum of SSD function

It is reasonable that in order to increase the sharpness of the SSSD function near the true solution, more emphasis should be given to reliable SSDs, while de-emphasizing unreliable ones. In other words, in general, not all the N images, used for the calculation of the SSSD in Eq. (6), make equal contributions to find the optimal S. Therefore, in this section, in order to reduce the ambiguities and improve the performance in determining the optimal solution for SSSD, let us define the weighted SSSD (WSSSD) function, in which each SSD is weighted, according to the degree of its reliability, given by

$$WSSSD(\mathbf{x}_0, S) = \frac{\sum_{i=1}^{N} w_i(\mathbf{x}_0) SSD_i(\mathbf{x}_0, S)}{\sum_{i=1}^{N} w_i(\mathbf{x}_0)} N, \tag{7}$$

where the weight $w_i(\mathbf{x}_0)$ for the ith SSD function is the measure of reliability, implying how confident the ith image is for the point $\mathbf{x}_0$ of the reference image. Similar to the lateral stereo case, we define the generalized baseline between the reference and target frames in general configuration as the distance from the point $\mathbf{C}_i$ to the line $\overline{\mathbf{C}_0\mathbf{x}_0}$.

$$w_i(\mathbf{x}_0) = \sin\theta_i \cdot |\mathbf{C}_i|, \tag{8}$$

where

$$\cos\theta_i = \frac{\mathbf{x}_0 \cdot \mathbf{C}_i}{|\mathbf{x}_0| \cdot |\mathbf{C}_i|}, \quad 0^\circ \leqslant \theta_i < 180^\circ. \tag{9}$$

Note that θ_i is the angle between the two vectors, $\mathbf{x}_0$ and $\mathbf{C}_i$. The generalized baseline is used for the reliability measure for the corresponding SSD function. In the case of lateral motion stereo where the camera motion is parallel to the image plane, as shown in Fig. 2(a), θ_i remains constant for all the image pairs. Thus, the weight $w_i(\mathbf{x}_0)$ is directly proportional to the length of the

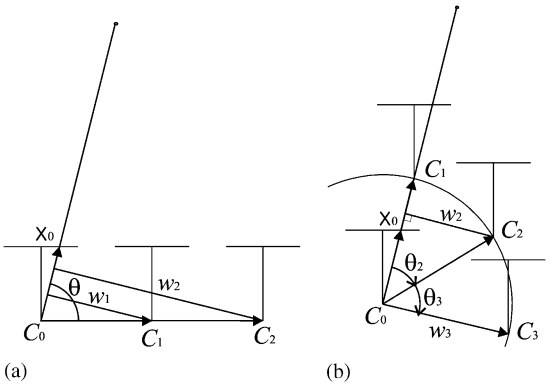

Fig. 2. Weight function.

baseline. This notion is based on the fact that a longer baseline generally yields smaller error [1]. For the case where $|\mathbf{C}_i|$ is constant, that is, all $\mathbf{C}_i$'s are on a surface of sphere of which center is $\mathbf{C}_0$, as shown in Fig. 2(b), $w_i(\mathbf{x}_0)$ is only proportional to $\sin\theta_i$. Note that when $\mathbf{C}_i$ is on the optical ray $\overline{\mathbf{C}_0\mathbf{x}_0}$, $w_i(\mathbf{x}_0)$ becomes zero, implying that this image is no longer reliable for the computation of the depth of point $\mathbf{x}_0$. For example, in the case of axial motion stereo, we cannot acquire any information from the feature point in the right center of an image. While, if $\overline{\mathbf{C}_0\mathbf{C}_i}$ is perpendicular to $\overline{\mathbf{C}_0\mathbf{x}_0}$, then the image can be considered to be the most reliable one.

The effectiveness of the proposed weighting scheme is verified using synthetic test images. Fig. 3(a) shows a reference image with a test feature point near the optical center, marked by '+', and Figs. 3(b) and (c) present the images taken under lateral and axial motions, respectively. The individual SSD functions for these images under lateral and axial motion are shown in Fig. 4(a). From Fig. 4(a), it is seen that the image under axial motion has more uncertainty in determining the depth of this feature point, while the image under lateral motion yields more reliable result. In Fig. 4(b), the SSSD and the WSSSD functions for these two image pairs are shown, indicating clearly that an ability of localizing the true estimate is significantly improved by the WSSSD scheme. The calculated weights for the first and second image are 14.14 and 2.64, respectively. The quantitative evaluation of two object functions for this test is provided in Table 1, in which the estimation error for the WSSSD is almost one seventh of that for the simple SSSD.

2.3. Elimination of occlusion problem using the selective WSSSD function

As the separation of the positions of the camera increases, usually two stereo images to be matched become

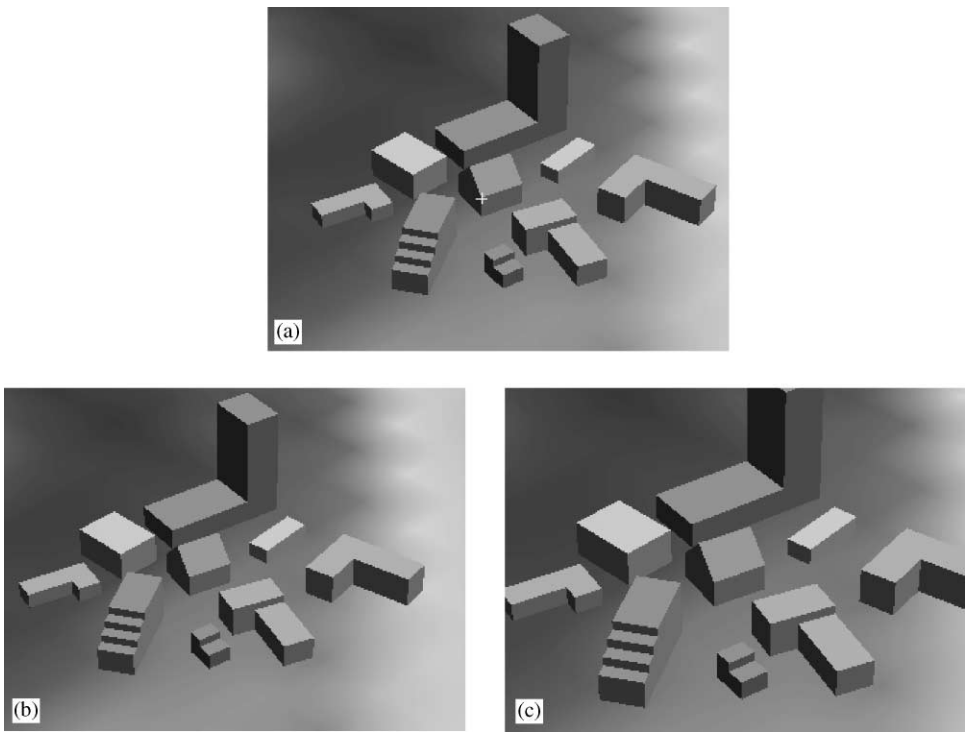

Fig. 3. Images for test of employing weight functions: (a) reference image, (b) first image (lateral motion), (c) second image (axial motion).

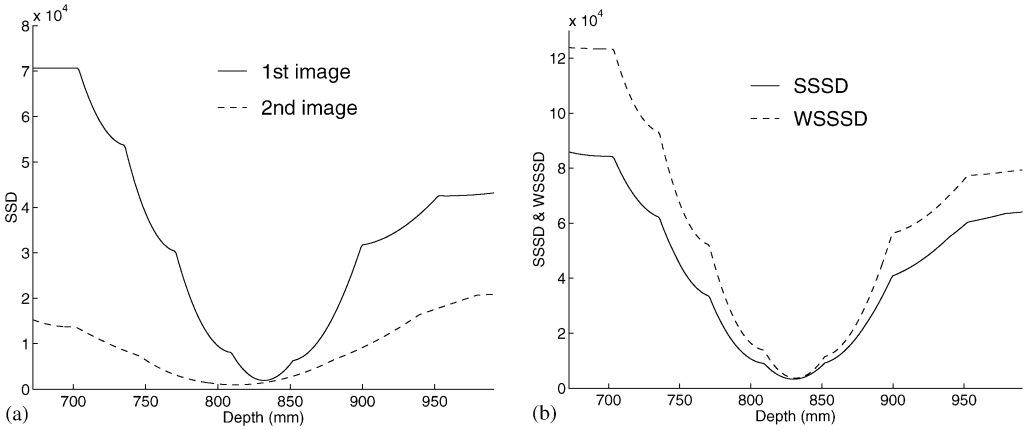

Fig. 4. Results on the test of employing weight functions: (a) SSDs, (b) SSSD and WSSSD.

Table 1
Reconstruction error of SSSD and WSSSD

	Object function	Error (mm)
(a)	SSSD	2.33 mm
(b)	WSSSD	0.33 mm

less similar. Thus, some objects obtained by one camera may not even be visible to the other [1]. Note that in the case of such occlusion or corrupted noise in image pairs, the corresponding SSD function may yield no local minimum near the true depth value, although it has relatively large weights. This situation tends to make the shape of the overall WSSSD function dull, making it very difficult to find the accurate estimate of the depth.

In this section, we propose an effective method to detect and exclude these undesirable SSD functions for computing Eq. (7), so that more reliable and accurate depth reconstruction can be achieved. Note that for an image without occlusion or noise, the SSD function tends to have a local minimum, close to the true depth value. Thus, it is probable that the depth estimate could be found within the region where many SSD functions have local minimums. The SSD functions that have no local minimum in this region should be excluded in evaluating the object function, because they cannot give any more reliable information in that region.

In order to find the region where many SSD functions have local minimums, we first obtain the initial individual depth estimates S's which locally minimize each single SSD function. In general, each SSD function could have single, multiple, or even no local minimum at all. If K local minimums are obtained from N SSD functions, by sorting these values by S, we can construct an ordered set of the positions of the local minimums, $\mathscr{A}$ as

$$\mathscr{A} = \{S_1, S_2, \ldots, S_K\}, \tag{10}$$

where

$$S_i \leqslant S_j \quad \text{if } i < j. \tag{11}$$

Then, let us define the blocking function, given by

$$d_{k,i} = \begin{cases} 1, & SSD_i(\mathbf{x}_0, \cdot) \text{ has local minimum in } [S_k, S_k + \beta], \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

which indicates whether the ith SSD function has local minimum near S_k in the right-hand side within a window of size β. And, then the sum of the blocking functions for all the SSD functions is defined by

$$N_k = \sum_{i=1}^{N} d_{k,i}, \tag{13}$$

which is the very number of SSD functions that have local minimums in $[S_k, S_k + \beta]$. If N_k is 1, only one SSD function has local minimum at S_k. Whereas, if N_k becomes N, all SSD functions have their local minimums in $[S_k, S_k + \beta]$. Thus, as N_k increases, it is more likely that the true depth value would be included in $[S_k, S_k + \beta]$. In our algorithm, the candidate region for the depth estimate is chosen if N_k is larger than $N/2$. It means that at least half of the images must provide reliable information about the feature points.

Now, using the blocking function and the sum in Eqs. (12) and (13), we can define the Selective WSSSD

(SWSSSD) function as follows.

$$SWSSSD_k(\mathbf{x}_0, S) = \frac{\sum_{i=1}^{N} d_{k,i} \cdot w_i(\mathbf{x}_0) \cdot SSD_i(\mathbf{x}_0, S)}{\sum_{i=1}^{N} d_{k,i} \cdot w_i(\mathbf{x}_0)} \cdot N_k, \tag{14}$$

where the blocking function is used for determining whether the ith SSD function would be included or not. Note that we only consider the $SWSSSD_k$ functions for the corresponding regions $[S_k, S_k + \beta]$ in which N_k is greater than $N/2$, and determine the depth candidates which minimize the object functions in each region. The final depth estimate is then obtained by choosing the depth value among the candidates which gives the global minimum. Note that although some feature points in the reference image might be invisible in some other images, due to the occlusion or noise, the depth of those feature points can be recovered robustly, as long as more than half of the images give reliable information for those points.

Let us consider an example in which one reference image and four other images are employed, and four SSD functions for each pair of images have the local minimums as follows:

$$SSD_1: S = 0.4, S = 1.1, S = 3.6,$$

$$SSD_2: S = 0.7, S = 2.5, S = 4.5,$$

$$SSD_3: S = 2.8, S = 5.2,$$

$$SSD_4: S = 3.0, S = 5.0. \tag{15}$$

Then, the set $\mathscr{A}$ is constructed as

$$\mathscr{A} = \{0.4, 0.7, 1.1, 2.5, 2.8, 3.0, 3.6, 4.5, 5.0, 5.2\}, \tag{16}$$

which is plotted in Fig. 5. As shown in this figure, by using the window size $\beta = 1$, the blocking functions $d_{k,i}$ and the sum of them N_k for each local minimum S_k are determined easily. Among these values, since only N_4, N_5, and N_8 are larger than the threshold $N/2(= 2)$, the corresponding functions $SWSSSD_4$, $SWSSSD_5$, and $SWSSSD_8$ are to be examined exclusively, and the final depth is determined as the one which gives the minimum value of those functions in each corresponding window region. Note that in evaluating each SWSSSD function, only three reliable SSD functions among four are employed according to the corresponding blocking function. That is, the other SSD function that does not give reliable information can be effectively excluded.

Note that as the size of the window β increases, more local minimums would be included in the same window, which may degrade the discrimination ability in selecting the reliable SSD functions. On the contrary, if the size of the window becomes too small, then the performance could be easily affected by noise. Empirically, we observed that except for the extreme cases of too large or too small, the influence of the size of β on the performance of algorithm is negligible.

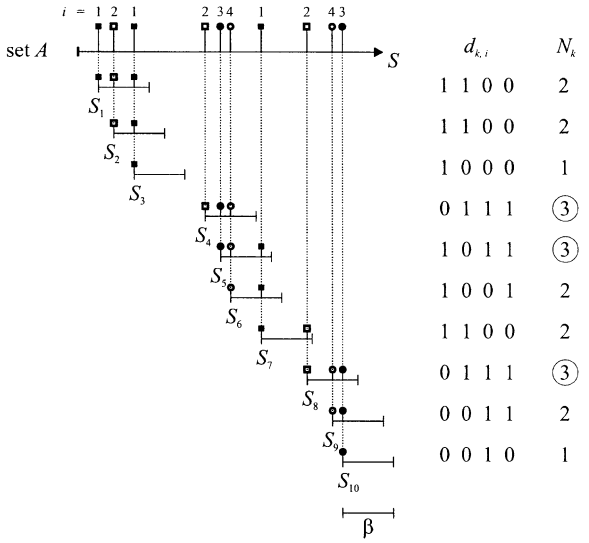

Fig. 5. Example of selection.

We examine the performance of the proposed method using several test images with occlusion. Fig. 6(a) shows a reference image with a test feature point, marked by '+', and Figs. 6(b)–(f) show the images taken under a general motion. Note that the feature point is invisible, due to occlusion by other parts, as shown in Figs. 6(b) and (f). The SSD functions for each test images are shown in Fig. 7(a). From Fig. 7(a), it is observed that the SSD functions, corresponding to the occluded images, do not have local minima near the true depth, while the others have. Fig. 7(b) shows the WSSSD and the SWSSSD functions for all the images. But, the two SSD functions for the occluded images in Figs. 6(b) and (f) are not included in calculating the SWSSSD. Note that the SWSSSD function yields a sharp minimum at the true depth, while the WSSSD yields a false estimate. The quantitative measures of the estimation error are provided in Table 2, demonstrating that the proposed SWSSSD function can recover the depth quite satisfactorily, even in the presence of occlusion or noise in images.

3. Experimental results

We test the proposed algorithm on several images, both synthetic and real. Fig. 8 shows the data set we have synthesized. The dimensions of the objects in the test images are in the range of 12–48 mm in each side, and the camera of the reference image is located about 789 mm away from the objects. Seventeen images are used for this

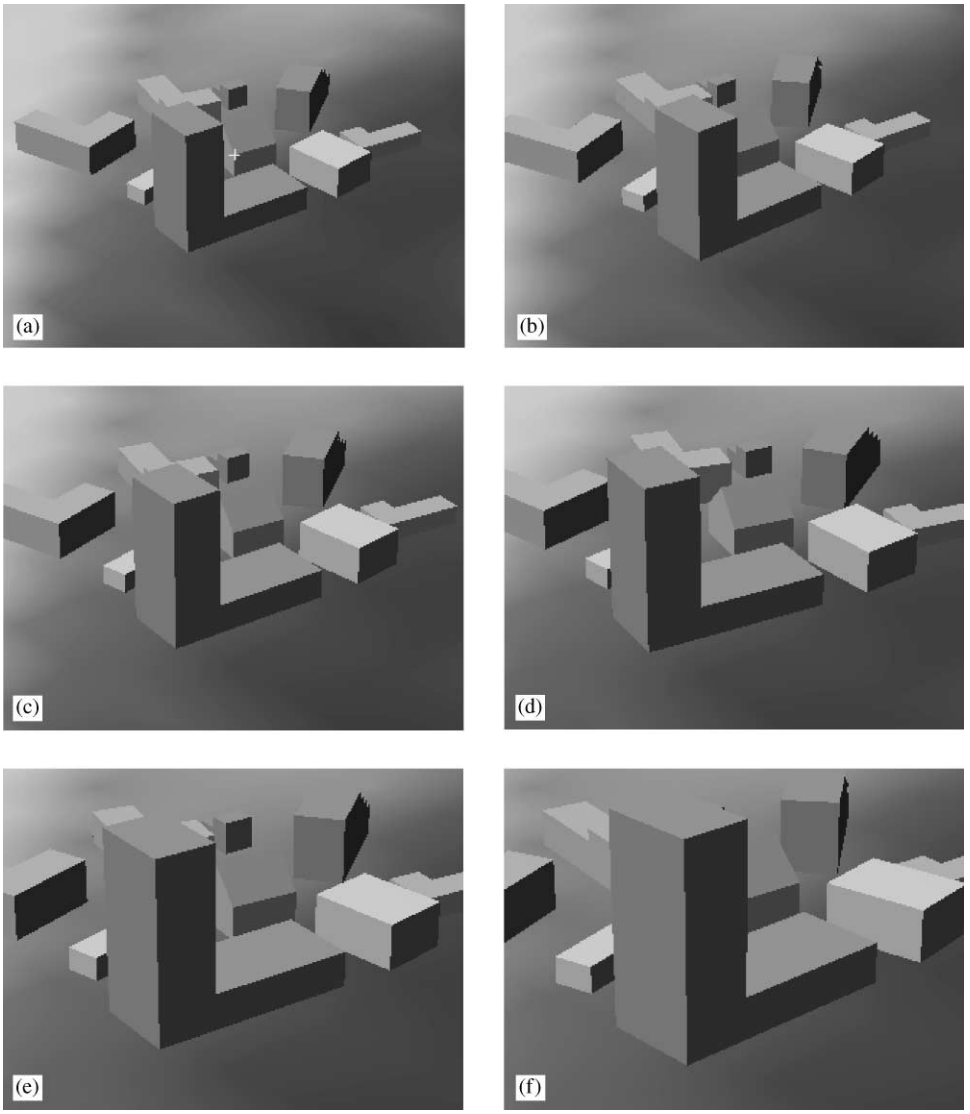

Fig. 6. Images for the test of image selection: (a) reference image, (b) first image (occlusion), (c) second image (no occlusion), (d) third image (no occlusion), (e) fourth image (no occlusion), (f) fifth image (occlusion).

test. Figs. 8(a) and (b) are the reference and the 16th image, and 102 corner points, marked in the reference image, are selected by hand for the feature points.

In Fig. 9, the average depth error for the 102 feature points is compared under different conditions. Fig. 9(a) shows the depth error, caused by using individual pair of images. It is noted that, in this case, the error is quite dependent on each image pair used. While, as shown in Fig. 9(b) in which multiple images are combined, the error decreases substantially as the number of images increases. We also examine the effects of the size of window W and the searching step ΔS, respectively. Figs. 9(c) and (d) show the average depth error, according to the window size and the search step size, respectively. From these results, it can be concluded that the window size of 5×5 provide the best performance. As the searching step becomes smaller, the error is reduced. However, when it reaches a certain threshold value, the error is converged. Thus, searching step is not important as long as it is below the threshold value. This threshold value can be varied to the distance from the camera to the object. In all our experiments, searching step of 1 mm is satisfactory. Also, in order to demonstrate the accuracy of the reconstruction qualitatively, the extracted 3-D

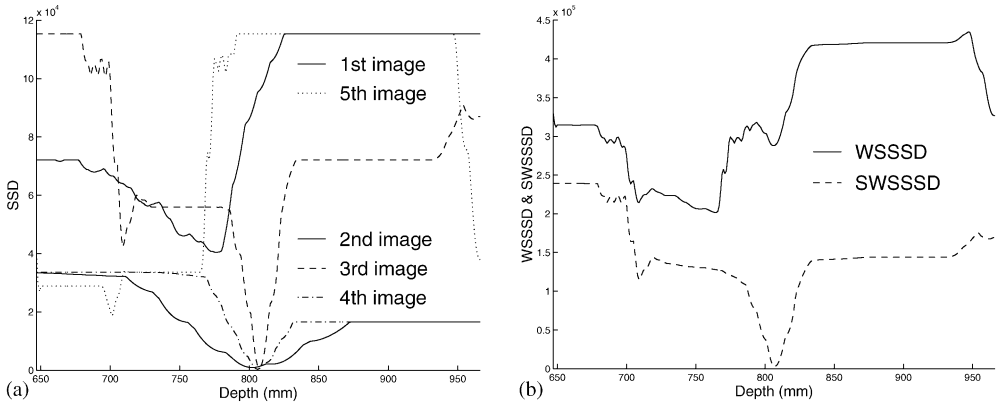

Fig. 7. Results on the test of image selection: (a) SSDs, (b) WSSSD and SWSSSD.

Table 2
Reconstruction error of WSSSD and SWSSSD

	Object function	Error (mm)
(a)	WSSSD	42.50
(b)	SWSSSD	0.50

information of the feature points and the connected lines are rendered in different viewing directions as shown in Fig. 10.

As presented in Table 3(c), the average depth reconstruction error for the 102 feature points using the SWSSSD function is found to be 0.82 mm. Considering the fact that the camera is about 789 mm away from the objects, 0.82 mm depth error can be considered as quite negligible. To compare the result with those of other object functions, we present the average error using the mean error of each binocular stereo matching, and that of the SSSD in Table 3(a) and (b), respectively. From these results, it is also observed that the proposed the SWSSSD algorithm reduces the error by 68%, compared to that of the SSSD.

Fig. 11 shows the real block images we test. The sizes of blocks are in the range of 30–50 mm in each side, and the initial location of the camera is approximately 1 m away from these blocks. Eight images are taken by moving the camera, and the reference and the 7th image are shown in Figs. 11(a) and (b), respectively. We select 29 feature points and 37 feature lines in the reference image, which are shown in Fig. 11(a). Note that for this test of real objects, true depth information for each feature point cannot be available. However, since we know the relative distances between the feature points, that is, the actual lengths of each block, we can utilize this information to measure the accuracy of the depth reconstruction. The average error for the 37 feature lines using the SWSSSD is presented in Table 4, along with those using the mean of the binocular stereo and the SSSD methods. Comparing the depth range of the objects from the camera, we can conclude that the reconstruction error is negligible. The extracted 3-D objects in different viewing directions are shown in Fig. 12.

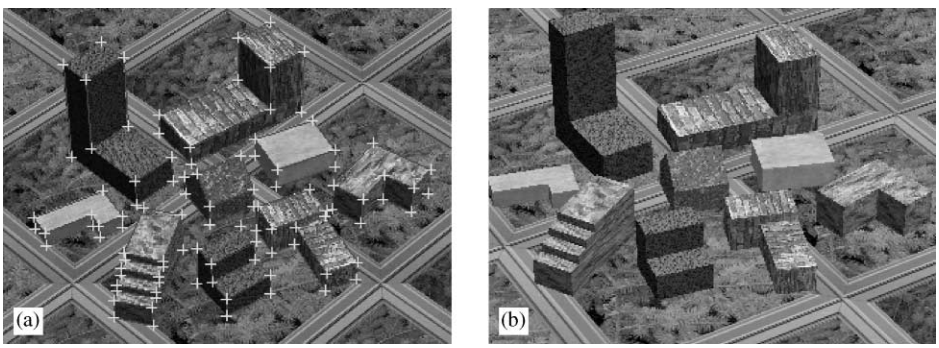

Fig. 8. Synthetic building images: (a) reference image, (b) 16th image.

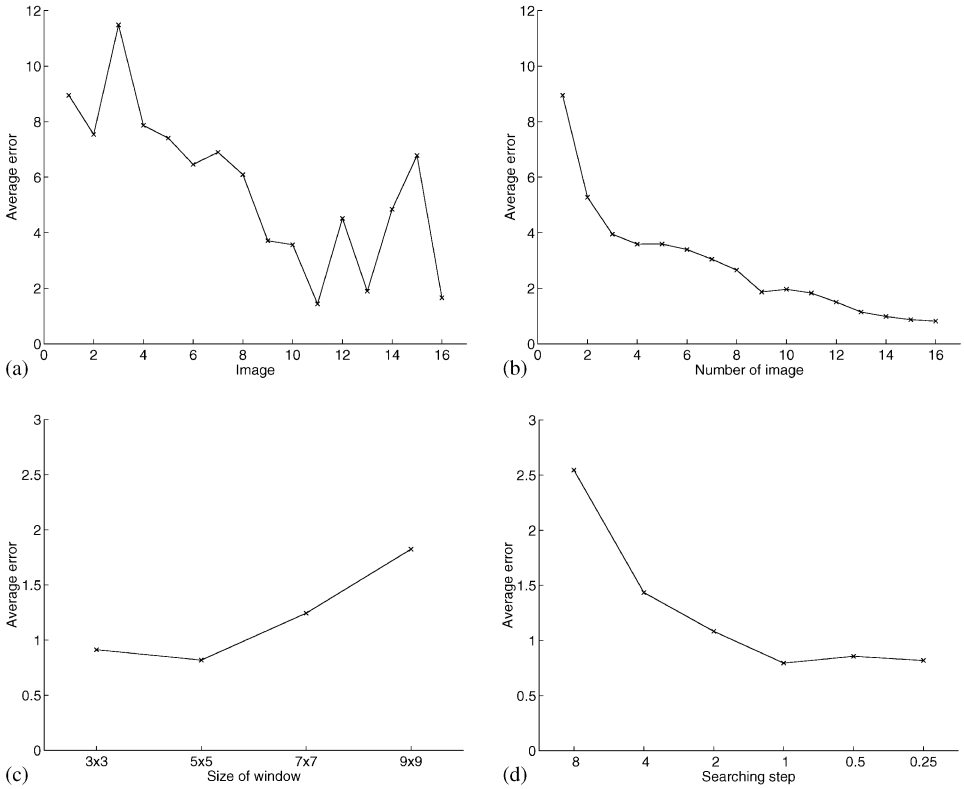

Fig. 9. Average error for the synthetic building images: (a) binocular stereo, (b) SWSSSD, (c) size of window, (d) searching step.

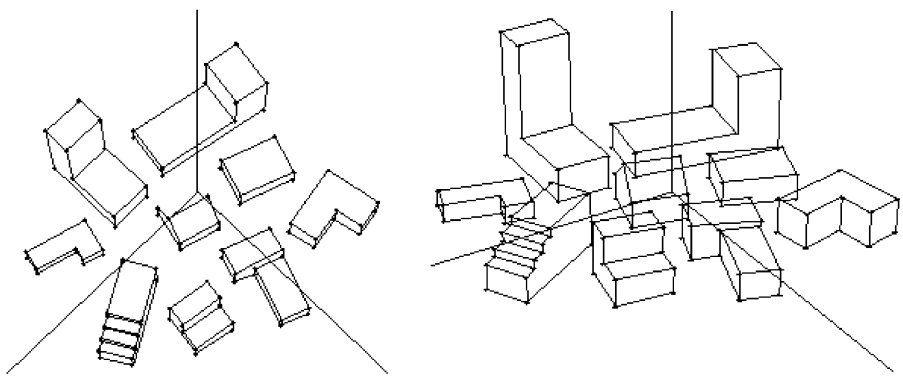

Fig. 10. Reconstructed buildings in different viewing directions.

Fig. 13 shows another test set of CAN images. Seven images are taken, and the reference image and the sixth image are presented in Figs. 13(a) and (b), respectively. Hundred feature points on the surface of the CAN are chosen for this test as shown in Fig. 13(a). Note that since the feature points are on the surface of the CAN of which actual radius is 50.5 mm, the reconstructed points also can be fitted to a cylindrical surface. The radius of the resulting fitted cylinder and the mean absolute error of the points to this cylinder for each method are calculated, which are provided in Table 5. Compared to the other two methods, the SWSSSD method yields the best result with the smallest error. The extracted 3-D feature points are presented in Fig. 14, showing that the reconstruction is qualitatively quite accurate.

Table 3
Average reconstruction error for the synthetic building images

	Object function	Error (mm)
(a)	Mean of binocular stereos	2.49
(b)	SSSD	2.58
(c)	SWSSSD	0.82

Table 4
Average reconstruction error of the block images

	Object function	Error (mm)
(a)	Mean of binocular stereos	13.87
(b)	SSSD	1.54
(c)	SWSSSD	1.19

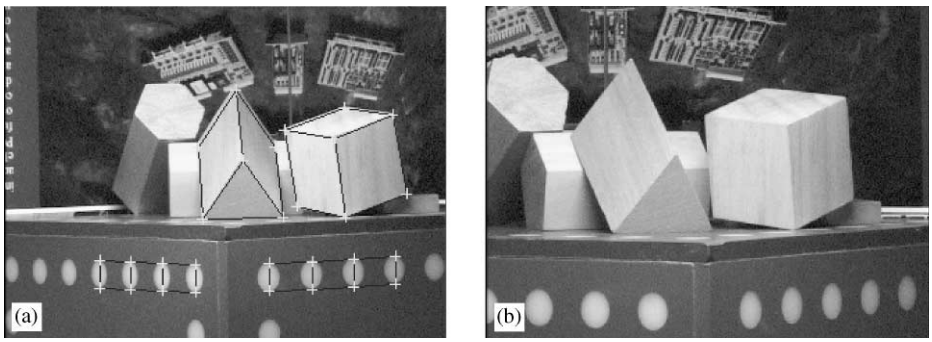

Fig. 11. Block images: (a) reference image, (b) seventh image.

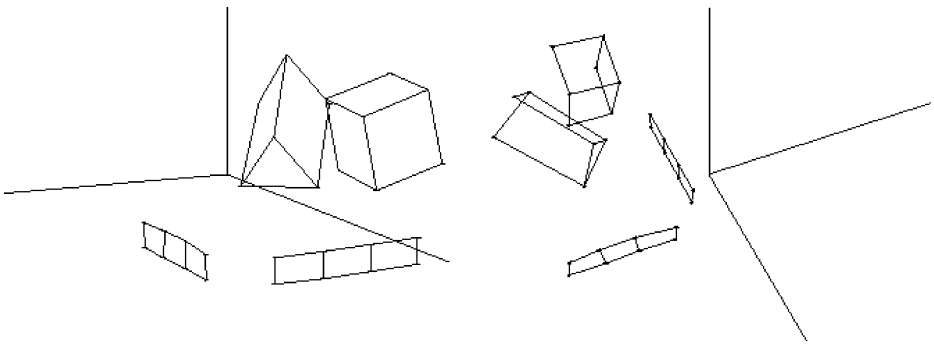

Fig. 12. Reconstructed blocks in different viewing directions.

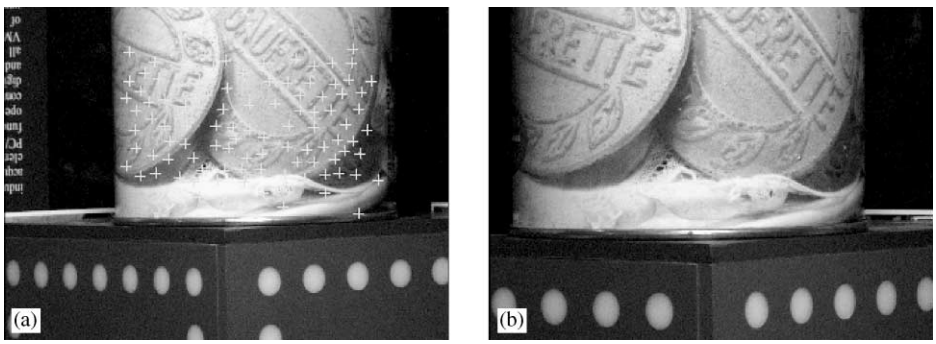

Fig. 13. CAN images: (a) reference image, (b) sixth image.

Table 5
Reconstructed radius and the mean of absolute error for the CAN images

	Object function	Radius (True: 50.5 mm) (mm)	Mean of absolute error (mm)
(a)	Mean of binocular stereos	52.67	3.60
(b)	SSSD	51.70	1.10
(c)	SWSSSD	51.20	0.93

The last test is performed using the calibrated imaging lab. (CIL) images, which are commonly used for the test of stereo algorithm [12]. Eleven images are used, and the reference and the 10th images are shown in Figs. 15(a) and (b), respectively. The camera is located about 1.8 m away from the buildings, of which the sizes are about 100–200 mm. Similarly, for the 26 feature points of which ground truths are given, the results are summarized in Table 6. These results also reveal that the proposed SWSSSD scheme greatly enhances the accuracy of the depth estimation.

Table 6
Average reconstruction error for the CIL images

	Object function	Error (mm)
(a)	Mean of binocular stereos	11.77
(b)	SSSD	4.78
(c)	SWSSSD	3.46

4. Conclusion

In this paper, we have proposed a new motion stereo algorithm to extract the 3-D information from multiple images, taken under a general motion. The object functions for individual stereo image pairs were represented,

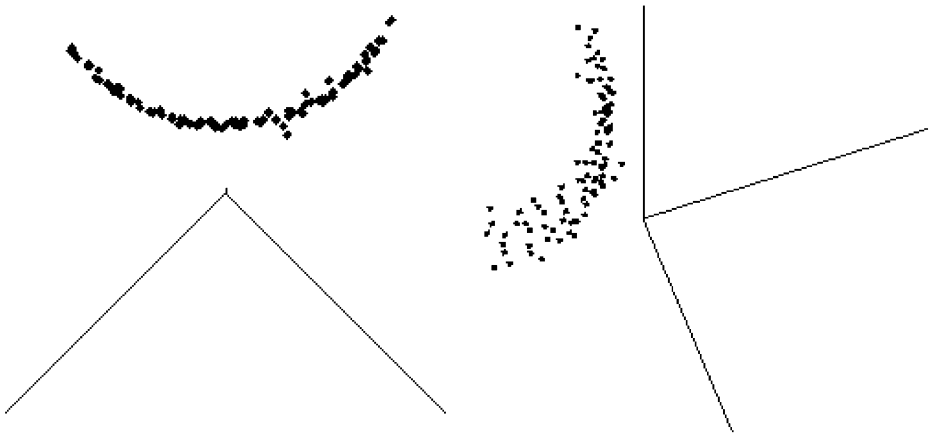

Fig. 14. Reconstructed CAN surface in different viewing directions.

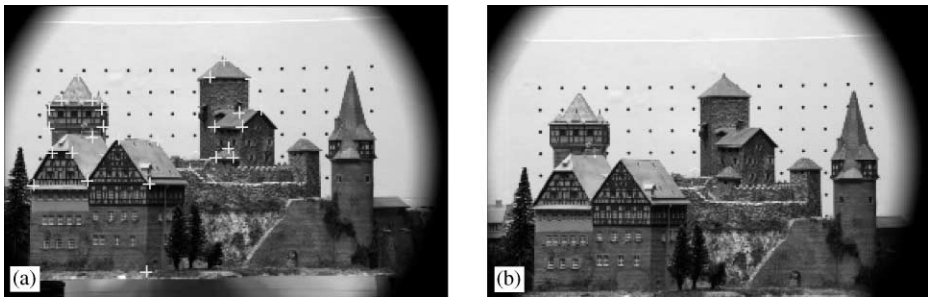

Fig. 15. CIL images: (a) reference image, (b) 10th image.

in terms of the depth, then these object functions were integrated in batch mode. By imposing a weight to each object function, according to the measured reliability, and employing the object functions of nonoccluded images selectively, the stability and the accuracy of the reconstruction could be improved significantly. Experiments on the synthetic and real stereo images demonstrated that the proposed algorithm provides quite satisfactory results. Since the proposed algorithm is based on a general camera motion, it can be easily applied to the real applications, such as obstacle avoidance and smart missile navigation.

References

[1] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, 1986.
[2] O. Faugeras, Three-Dimensional Computer Vision, MIT Press, Cambridge, MA, 1993.
[3] N. Alvertos, D. Brazakovic, R.C. Gonzalez, Camera geometries for image matching in 3-D machine vision, IEEE Trans. Pattern Anal. Mach. Intell. 11 (9) (1989) 897–915.
[4] V. Rodin, A. Ayache, Axial stereovision: Modelization and comparison between two calibration methods, Proceedings of the International Conference on Image Processing, Vol. 2, November 1994, pp. 725–729.
[5] R.Y. Tsai, Multiframe image point matching and 3-d surface reconstruction, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-5 (2) (1983) 159–174.
[6] M. Okutomi, T. Kanade, A multiple-baseline stereo, IEEE Trans. Pattern Anal. Mach. Intell. 15 (4) (1993) 353–363.
[7] T. Kanade, A stereo machine for video-rate dense depth mapping and its new applications, Proceedings of the Computer Vision and Pattern Recognition, June, 1996, pp. 196–202.
[8] R. Collins, A space-sweep approach to true multi-image matching, Proceedings of the Image Understanding Workshop, Vol. 2, February 1996, pp. 1213–1220.
[9] P. Pritchett, A. Zisserman, Wide baseline stereo matching, Proceedings of the International Conference on Computer Vision, January 1988, pp. 754–760.
[10] D.W. Murray, Recovering range using virtual multicamera stereo, Comput. Vision Image Understanding 61 (2) (1995) 285–291.
[11] J.R. Stenstrom, C.I. Connolly, Constructing object models from multiple images, Int. J. Comput. Vision 9 (3) (1992) 185–212.
[12] Calibrated Imaging Laboratory at Carnegie Mellon Univerity, CIL Stereo Dataset, http://www.cs.cmu.edu/~cil/cil-ster.html.

About the Author—JA SEONG KU received B.S. and M.S. degrees in Electrical Engineering from Seoul National University, Seoul, Korea in 1993 and 1995, respectively. He is currently working towards Ph.D. degree at Seoul National University. His research interests are in digital signal processing, including image processing and computer vision.

About the Author—KYOUNG MU LEE received B.S. and M.S. Degrees in Control and Instrumentation Engineering from Seoul National University, Seoul, Korea in 1984 and 1986, respectively, and Ph. D. degree in Electrical Engineering from the USC (University of Southern California), Los Angeles, California in 1993. From 1993 to 1994, he was a research associate in the SIPI (Signal and Image Processing Institute) at USC, and was with the Samsung Electronics Co., Ltd. in Korea as a senior researcher from 1994 to 1995. Since August 1995, he has been an assistant professor in the department of Electronics and Electrical Eng. of the Hong-Ik University. His primary research interests include computational vision, image understanding, pattern recognition, man–machine interface and multimedia applications.

About the Author—SANG UK LEE received the B.S. degree from Seoul National University in 1973, the M.S. degree from Iowa State University in 1976, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1980, all in Electrical Engineering. In 1980, he was with General Electric Company, Lynchburg, VA, and in 1981–1983, he was a Member of Technical Staff, M/A-COM Research Center, Rockville, MD. In 1983, he joined the school of Electrical Engineering at Seoul National University as an Assistant Professor, where he is now a Professor. His current research interests are in the areas of image processing, digital communication, and computer vision. Dr. Lee is a member of Phi Kappa Phi.