IEICE TRANS. INF. & SYST., VOL.E91–D, NO.1 JANUARY 2008

PAPER

Key-Frame Selection and an LMedS-Based Approach to Structure and Motion Recovery

Yongho HWANG[†], *Student Member*, Jungkak SEO[†], *and* Hyunki HONG[†*a)], *Nonmembers*

SUMMARY Auto-calibration for structure and motion recovery can be used for match move where the goal is to insert synthetic 3D objects into real scenes and create views as if they were part of the real scene. However, most auto-calibration methods for multi-views utilize bundle adjustment with non-linear optimization, which requires a very good starting approximation. We propose a novel key-frame selection measurement and LMedS (Least Median of Square)-based approach to estimate scene structure and motion from image sequences captured with a hand-held camera. First, we select key-frames considering the ratio of number of correspondences and feature points, the homography error and the distribution of corresponding points in the image. Then, by using LMedS, we reject erroneous frames among the key-frames in absolute quadric estimation. Simulation results demonstrated that the proposed method can select suitable key-frames efficiently and achieve more precise camera pose estimation without non-linear optimization.

key words: *auto-calibration, key-frames selection, corresponding points, absolute quadric estimation, least median of square*

1. Introduction

Computer vision techniques have been applied for visual effects since the 1990's, and match move is one of the representative research areas. This makes it possible to insert synthetic 3D objects into real, but un-modeled scenes, and create views from given camera positions so that they appear to move as if they were part of the real scene [1]. For stable 3D appearance changes of the object from the camera, camera pose estimation is needed. At the same time, the 3D structure of the scene is used for placement of the objects with respect to the real scene free of occlusion. Manually compositing a synthetic object on real scenes is a very difficult and time consuming process that might take days or weeks. In order to automate this process, we introduce reliable camera pose and scene geometry recovery that works with auto-calibration.

Multi-view pose and geometry analysis has attracted much attention in recent years [2]–[4]. Auto-calibration is the process of determining camera parameters directly from multiple images in a sequence obtained by a hand-held camera. Projective reconstruction is necessary as a preceding step in auto-calibration for scene structure and motion estimation, and it is classified into merging-based and factorization methods. Sawhney presented a method that accurately estimates relative camera pose from the fundamental matrix over a video sequence [1]. Pollefeys proposed a 3D modeling technique over image sequences from a hand-held camera, and then extended that for AR-systems [4], [5]. Gibson described an improved feature-tracking algorithm based on the KLT (Kanade-Lucas-Tomasi) tracker, and presented a robust hierarchical scheme merging sub-sequences together to form a complete projective reconstruction [6]. Sturm presented a factorization method to calculate all camera projection matrices and structure at the same time, which suffers from less drift and error accumulation [7]. The drawback of factorization methods relying on matrix decomposition is that all corresponding points must remain in all views from the first frame to the last. The merging-based projective reconstruction method is able to solve this problem [8], [9]. Sequential merging algorithms are heavily dependent on a good initial estimate of structure, and are susceptible to drift over long sequences, so the resulting error accumulates over time. Hierarchical merging algorithms were proposed to improve sequential methods, and have the advantage of the drift error being more evenly distributed over the entire sequence [6]. In our experiments, we evaluate the accuracy of the proposed auto-calibration algorithm on the sequential and hierarchical approach.

In general, motion between frames has to be fairly small so that a precise correspondence can be established by using automatic matching, while significant parallax and a large baseline is desirable for obtaining a well-conditioned problem [10]. A good choice of the frame from an image sequence can produce more appropriate input for pose and geometry recovery, thereby improving the final result. Hence the goal of key-frame selection is to select a minimal subsequence of feature views from the images, such that correspondence matching still works for all pairs of adjacent frames in the sub-sequence. Sawhney mentioned that the key-frame may be chosen using the following criteria: when parallax motion is beyond a certain threshold and there is a change in the number of feature tracks [1]. Nister presented a frame decimation scheme based on global motion estimation between frames and a sharpness measure, which is the mean square of the horizontal and vertical derivatives, evaluated as finite differences [10]. In this approach, computational time is mainly dependent on the image size, because the image sharpness has to be evaluated. Gibson's measure includes the fraction of features that were reconstructed in

Manuscript received May 18, 2007.
Manuscript revised September 7, 2007.

[†]The authors are with the Computer Graphics & Media Laboratory, Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul 156–756, Korea.
*Correspondence author.
a) E-mail: honghk@cau.ac.kr

DOI: 10.1093/ietisy/e91–d.1.114

the previous key-frame pair, but cannot possibly be reconstructed in this pair [6]. In addition, the median epipolar error between every two views is estimated, which is a very time consuming process.

In this paper, we propose a novel measure for key-frame selection and an LMedS-based approach to projective reconstruction. Our quantitative measure takes into account the ratio of the number of corresponding points and feature points, the homography error and the distribution of corresponding points in the image. Carefully selecting key-frames enables camera pose and scene geometry recovery, which is a relatively expensive process, to be performed on a smaller number of views. In addition, video sequences with different amounts of motion per frame become more isotropic after frame decimation. Then we reject an erroneous frame among key-frames causing the absolute quadric estimation to fail by using LMedS (Least Median of Square). The LMedS algorithm chooses among the entire tested hypothesis the one that has the least median squared residual on the entire absolute quadric sets. The absolute quadric is re-estimated from the selected camera matrix set, and we recover the camera matrices of the rejected frames by the camera resection [11], [12]. We can obtain projective reconstruction by decomposing the absolute quadric. Finally, we determine a rectifying homography and transform the projective to a metric reconstruction [3]. In Fig. 1, the shaded regions compare contributions with the previous method. We evaluated various projective reconstruction methods and embodied an automatic match move to insert synthetic 3D objects into real scenes and create their views from the recovered camera positions.

This paper is laid out in the following format: Section 2 describes auto-calibration using absolute quadric, and Sect. 3 discusses key-frame selection. After details of our LMedS based on absolute quadric estimation are given in Sect. 4, we tackle the experimental results for synthetic and real scenes in Sect. 5. Finally, Sect. 6 discusses the conclusion.

2. Auto-Calibration

In this section, we briefly review the essence of the auto-calibration method. Given point matches from more than two images, projective structure and motion can be computed without camera parameters. In order to upgrade the projective structure to metric reconstruction, traditional methods first calibrate a camera by using an object with known 3D Euclidean geometry and a calibration pattern. Then, a metric structure of the given scene can be acquired from the correspondences between images. Recently there has been an active research on auto-calibration algorithms to avoid setting the calibration box in the scene because pre-procedures for calibration have a number of limitations and involve the setting of equipment [3], [12].

In general, we obtain projective reconstruction from a set of images acquired by a camera with fixed internal parameters. Then we compute a rectifying homography H from auto-calibration constraints, and transform to this a metric reconstruction [12]. The process of projection (P)

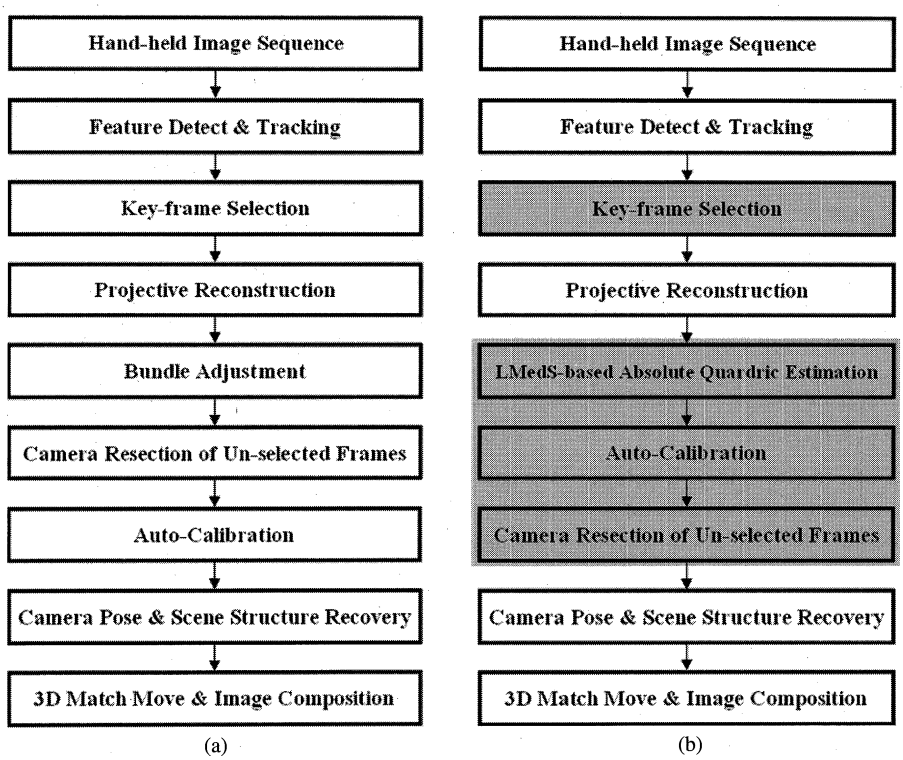


Fig. 1 Block diagram, (a) previous method, (b) proposed method.

of a point X in 3D to a point x in the image plane can be represented as follows:

$$x = PX = KR\left[I|-\tilde{C}\right]X = K[R|t]X,$$
$$K = \begin{bmatrix} f_x & s & x_o \\ & f_y & y_o \\ & & 1 \end{bmatrix} \quad (1)$$

where K is the camera calibration matrix with intrinsic parameters: focal length (f_x, f_y) in the x- and y-axis, skew factor (s), and principal point (x_o, y_o). I, $\tilde{C}$, t, and R are an identity matrix, the coordinates of the camera center in the world coordinate frame, $-R\tilde{C}$, and 3×3 rotation matrix representing the orientation of the camera coordinate frame, respectively.

We choose the world frame to coincide with the first camera, $P^1 = K[I|\mathbf{0}]$, so that $R^1 = I$ and $t^1 = \mathbf{0}$, where the superscript "1" represents the first frame. Because of the absence of intrinsic parameters, we eliminate K from projection matrices by using a linear transformation H as follows:

$$x_1 = P^1X = K[I|\mathbf{0}]X = K[I|\mathbf{0}]HH^{-1}X = [I|\mathbf{0}]X_p$$
$$x_2 = P^2X = K[R|t]X = K[R|t]HH^{-1}X = [M|p_4]X_p \quad (2)$$

where x_1 and x_2 are projections of two images of 3D point X. X_p is a projective reconstruction of X, where the subscript "p" means the projective reconstruction. M and p_4 represent KR and the last column of P^2, respectively. In the projective reconstruction, the camera position for the first view is established so that $P_p^1 = [I|\mathbf{0}]$.

We can reconstruct a scene up to a projective transformation by using only the corresponding points on the images. Calibration is the process of finding the transformation H, which can be obtained by decomposing the absolute quadric. In projective reconstruction, the dual image of the absolute conic (ω^*) is the projection of the absolute quadric (Ω_p) as follows:

$$\omega^* = P_p\Omega_pP_p^T = KK^T \quad (3)$$

When the camera has zero-skew, unit aspect ratio and the known principle point, the linear equations on Ω_p are generated from the zero entries in (3) as follows:

$$(P_p\Omega_pP_p^T)_{12} = (P_p\Omega_pP_p^T)_{13} = (P_p\Omega_pP_p^T)_{23} = 0,$$
$$(P_p\Omega_pP_p^T)_{11} = (P_p\Omega_pP_p^T)_{22} \quad (4)$$

where $(\,)_{ij}$ is element of i-th row and j-th column. We can estimate the absolute quadric from at least three images by using (4). Then we decompose the absolute quadric by using the eigen value decomposition, EVD, as

$$EVD(\Omega_p) = UDU^T = U\sqrt{D}\Omega_{\text{euc}}\sqrt{D}U^T = H\Omega_{\text{euc}}H^T \quad (5)$$

where Ω_{euc}, U and D are the absolute quadric in metric coordinate frame, an orthogonal matrix and a diagonal matrix, respectively, and the 0 eigenvalue in D is replaced by 1. Finally, from (5) we obtain metric camera motion and structure by applying H to projective coordinate frame [3].

3. Key-Frame Selection Measurement

In this section, we introduce the measures for key-frame selection. In order to achieve a suitable key-frame selection, we propose a new quantitative measurement that includes three factors: (i) the ratio of the number of corresponding points and feature points, (ii) the distribution of corresponding points about the frame and (iii) the homography error. These measurements are combined as follows:

$$S = w_1(1 - \frac{N_c}{N_F}) + w_2\sigma_c + w_3\frac{1}{H_{\text{err}}} \quad (6)$$

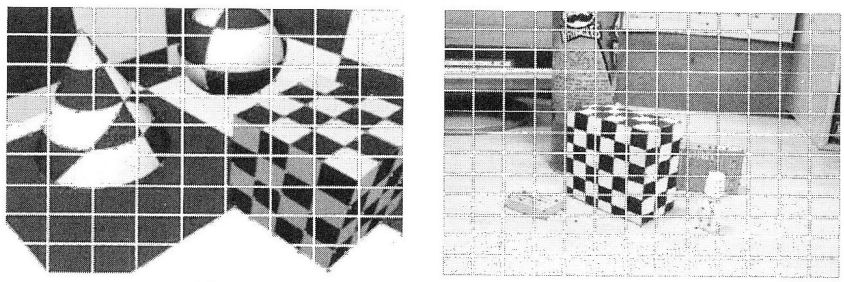
where S is the score to select the key-frame, N_c and N_F are the number of corresponding points and feature points. σ_c is the standard deviation of the density of corresponding points, and H_{err} is the homography error. $w_i(i = 1, 2, 3)$ is the weight used to alter the relative significance of each score. In experiment on various sequences, we determine three weights of the measure as: $w_1 = 3$, $w_2 = 1$, and $w_3 = 10$.

Correspondences between the first frame and successive frames gradually diminish as the frame number grows over the image sequence. The first term of (6) examines how many corresponding points remain on the successive frames. When the corresponding points are evenly distributed on the image, we can obtain a more precise fundamental matrix [13]. Since the fundamental matrix contains all available information on camera motion, more evenly distributed correspondences allow the improvement of 3D estimation results. The proposed method uses the standard deviation of the point density representing the distribution of corresponding points. In order to evaluate the degree of the point distribution on the image, we divide the entire image uniformly into subregions according to the number of the corresponding points, and calculate the point density of the subregion and that of the image [14]. Figure 2 shows the segmented sub-regions and corresponding points, which are represented by red points. The standard deviation can be represented as follows:

$$\sigma_c = \sqrt{\frac{1}{N_S}\sum_{i=1}^{N_S}\left(N_{Ci} - \frac{N_C}{N_S}\right)} \quad (7)$$

where N_S and N_C are the total number of subregions and corresponding points, and N_{Ci} is the number of corresponding points in i-th subregion, respectively.

The homography error is the median re-projection error when a planar projective homography is used for establishing corresponding points. Therefore, the homography error represents how many correspondences are distributed on a planar surface. If corresponding points are distributed on various surfaces, it is difficult to establish a one-to-one correspondence due to self-occlusion, increasing the homography error. This means that the homography error represents how much a camera moves between frames, and is used to evaluate a baseline length between two views. For these



(a) (b)

Fig. 2 Segmented sub-regions, (a) synthetic image, (b) real image.

reasons, Gibson used the homography error to select key-frames for sub-sequence grouping [6]. In addition, in order to estimate the fundamental matrix precisely, most corresponding points should not be placed on a planar surface or a line segment.

We determine the 2D homography matrix H such that $x_i' = Hx_i$ from a set of four 2D to 2D point correspondences $\{x_i \leftrightarrow x_i'\}$. This equation may be represented in terms of the vector cross product as $x_i' \times Hx_i = 0$, and we can obtain H from the simple linear equation using the singular value decomposition (SVD). In the case of more than four points, we use the normalized direct linear transform (DLT) [12]. Then we evaluate how many corresponding points on two frames satisfy the estimated homography as follows:

$$H_{\text{err}} = \frac{1}{N}\sum_{i=1}^{N} d(x_i', Hx_i) \tag{8}$$

where N is the number of correspondences between two frames, and $d(\ ,\ )$ is the Euclidean distance measurement.

After choosing the first frame as the key-frame, we examine all possible pairs of the first frame with consecutive frames in the sequence. Assuming that the key-frame has already been placed at the present frame, we achieve key-frame selection by evaluating the score for a pair of the current frame with the subsequent frame. This process continues until the ratio of the number of corresponding points and that of feature points is under 50 %. The frame with the lowest score of Eq. (6) is then marked as the key-frame within the sub-sequence. Therefore, because the key-frames are selected automatically from the sequence, the proposed algorithm need not adjust the number of the key-frame in advance as Gibson's [6] method does.

4. LMedS-Based Absolute Quadric Estimation

LMedS estimation scores the model by the median of the distances to all points in the data. Minimum sizes of subset samples are selected randomly with the number of samples obtained [12]. We can estimate absolute quadric by using (4) from at least three images. For more precise absolute quadric estimation, we propose a novel method using LMedS-based random sampling. We select random sets of projection matrices from the key frames, and derive the linear equations through (4). We automatically reject the frame with large errors among the key-frames, causing the absolute quadric estimation to fail. The obtained absolute quadric is projected to each camera matrix, and each residual is computed as follows:

$$r_i = \left(P_p^1 \Omega_p P_p^{1\,T} - P_p^i \Omega_p P_p^{iT}\right)_{\text{norm}} \tag{9}$$

where superscripts "1" and "i" represent the first frame and the i-th, respectively. Hence, P_p^1 is the first camera matrix that is an initial projection matrix [I|0] in the projective. We iterate the sampling and computing residuals, and then find the absolute quadric with the minimum median residual. From the minimum residual, a threshold for rejecting camera matrix with numerous errors is computed as follows [11]:

$$\tau = 2.5 \times 1.4826\,[1 + 5/(n - s)]\ \sqrt{r_{\text{median}}} \tag{10}$$

where r_{median} and n are the minimum median residual and the number of the camera matrices -1, and s is the number of the sampled camera ($s = 2$), respectively. The reader is referred to [15] for an explanation on these magic numbers, 2.5 and 1.4826. The advantage of LMedS-based estimation is that it requires no setting of thresholds or a priori knowledge of the variance of the error. The limitation is that it fails if more than half the data is outlying, for then the median distance will be to an outlier [12].

Reference [12] presented the equation to determine how many samples should be selected. The number of samples N is chosen sufficiently high to ensure with a probability, p, that at least one of the random samples of s points is free from outliers. Usually p is chosen at 0.99. Suppose w is the probability that any selected data point is an inlier, and thus $\varepsilon = 1 - w$ is the probability that it is an outlier. Then at least N selections (each of s points) are required, where $(1 - w^s)^N = 1 - p$, so that

$$N = \log(1 - p)/\log(1 - (1 - \varepsilon)^s). \tag{11}$$

The LMedS-based absolute quadric estimation is summarized as follows.

Step 1: Projective reconstruction process.
Step 2: Random sampling of two camera matrices except the first camera matrix.
Step 3: Estimate absolute quadric by (4) and compute the residual of each camera matrix by (9).

Step 4: Repeat 2-3, and find the absolute quadric with the minimum median residual.

Step 5: Reject erroneous camera matrices (frames) using the threshold from (10).

Step 6: Re-estimate the absolute quadric from the inlier camera matrix set.

Camera matrices of the rejected frames are recovered by the camera resection, and finally, the scene structred is reconstructed.

5. Experimental Results

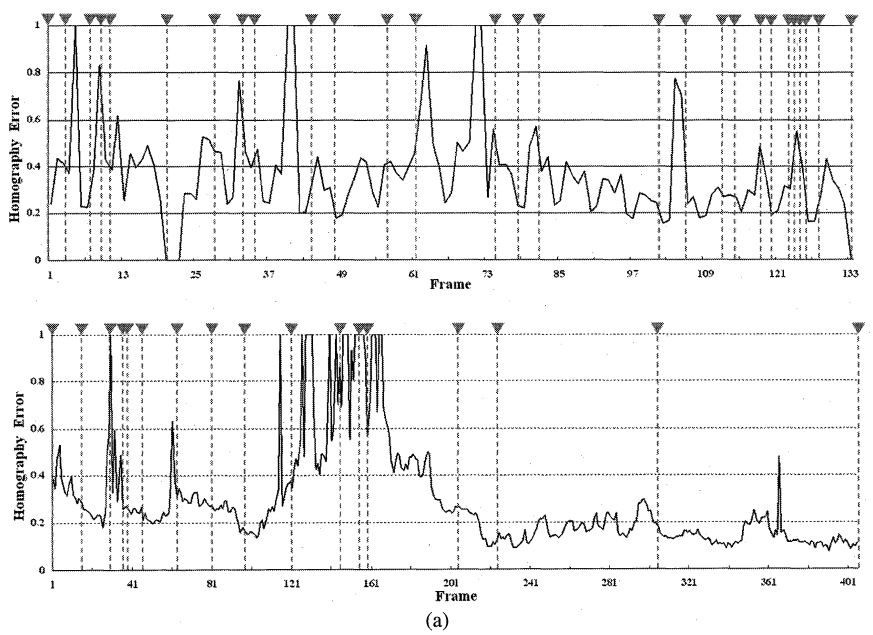
In order to evaluate the performance of the proposed key-frame selection algorithm, we have compared its results on 4 image sequences (Figs. 7 and 9) with those obtained by Nister's (A) and Gibson's (B) methods [6], [10]. Our simulation was performed on a PC using an Intel Pentium 4 2.3 GHz processor with 1 GB RAM. Table 1 shows the number of key-frames and the computation time. In the results obtained using Nister's method, computation time largely dependent on image size because image sharpness is evaluated. Gibson's method estimates the median epipolar error between every two views, and selects a relatively smaller number of key-frames than Nister's. However, Gibson's computation time is approximately the same as Nister's, because the estimation of a precise fundamental matrix and the median epipolar error is a very time consuming process. Instead of estimating the fundamental matrix and the median epipolar error, we consider the distribution of corresponding points on the frame, so the proposed algorithm (C) proves to be faster than previous methods.

In Fig. 3, the symbol "▼" represents positions of key-frames by three methods on the homography error of the "Fountain" and "Desk" sequences. The position and number of key-frames using the proposed method are almost similar to those of Gibson. These figures show that our proposed method enables key-frames to be more evenly distributed over the sequence. Because camera pose and scene geometry are estimated on the key-frames, a selection of fewer and more precise key-frames guarantees computational efficiency and accuracy. Therefore, though LMedS is a time consuming technique, the computation load of the proposed algorithm can be lessened.

Figure 4 shows the computed values of three terms in Eq. (6) with relative weights ($w_1 = 3$, $w_2 = 1$, and $w_3 = 10$) on "Desk_2" sequence. The previous method [6] gives more weight towards the homography error than other two factors,

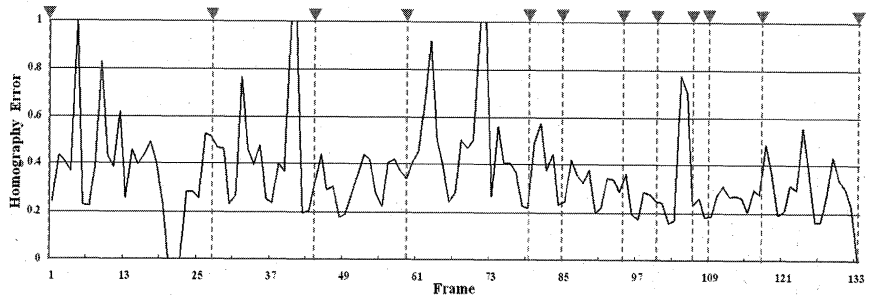
Table 1 Number of selected key-frame and computation time.

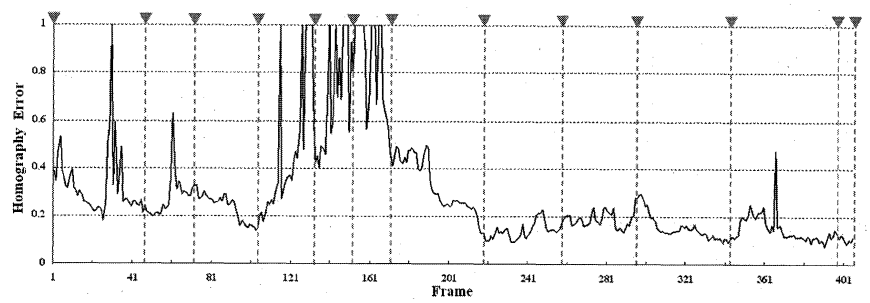
Image sequences			Number of the key frames			Computation time (sec)		
Type	Number of frames	Size of frame	A	B	C	A	B	C
Box	621	720 × 480	23	14	14	412	416	179
Desk	407	720 × 480	17	13	13	201	200	76
Fountain	134	320 × 240	28	12	12	16	38	10
Cottage	100	720 × 576	27	27	24	74	102	42



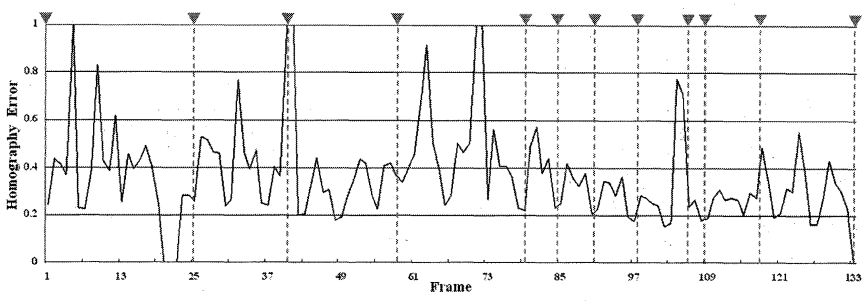
(a)

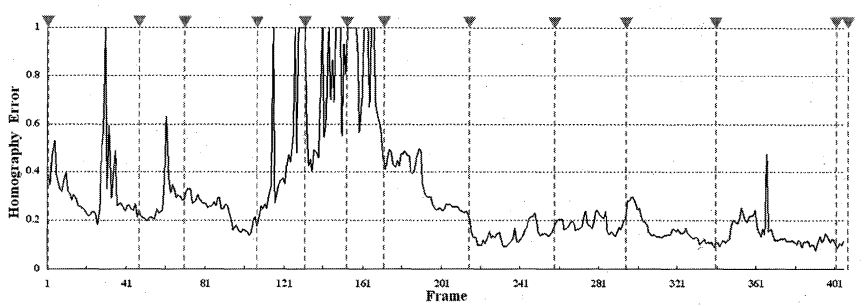
Fig. 3 Selected key-frames on the homography error of Fountain (left) and Desk (right), (a) Nister's method (A), (b) Gibson's method (B), (c) the proposed method (C).





(b)





(c)

Fig. 3 (Continued)

and takes into account the number of corresponding features in tracking process. The proposed algorithm includes not only these two items but also the measure of the distribution of correspondences in the present frame, instead of using the median epipolar error.

In our key-frame measure, the first term divides the image sequence into sub-sequences and the other two terms play an important role to determine the key-frame position. More specifically, the lower values in the second and the third term represent that most of corresponding points are more uniformly distributed and the camera positions moved more far away from the previous key-frame, respectively. Therefore, Fig. 4 shows that these results of two terms are much lower at the key-frame. In addition, the second term can alleviate the oulier effects in the homography term. Table 1 shows that the proposed method obtains the same results as those by Gibson [6], with better computation performance.

At first, we compared the auto-calibration algorithms on the synthetic data (Fig. 5) in terms of the estimated intrinsic parameters and reconstruction results. The camera is rotated around the model and moved along positive y-axis at the same time. The intrinsic parameters are fixed, and Gaussian noise is added to the input images. We have esti-

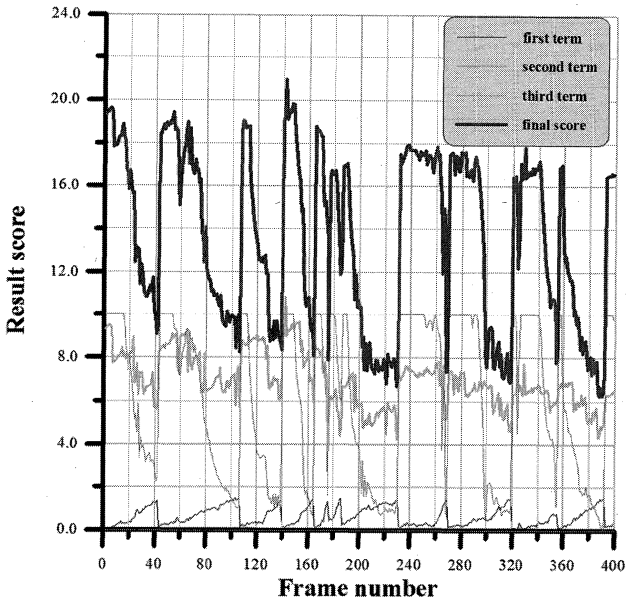


Fig. 4 Results of three terms in the selection measure with weight values.

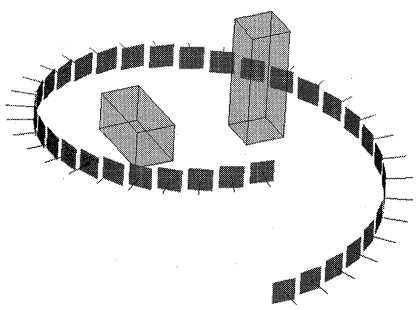


Fig. 5 Synthetic model and camera pose.

mated the absolute quadric with the linear method, bundle adjustment [12] and the proposed algorithm on two merging approaches: sequential merging [7] and a hierarchical merging [6]. Figures 6 and 7 represent a cumulative error of the camera's intrinsic parameters, and comparison of the camera pose recoveries. The sequential merging algorithm depends on an initial estimate of the structure, and the error propagates increasingly over time. On the other hand, since the hierarchical merging algorithm enables the error to propagate evenly over an entire sequence, the performance of the hierarchical merging algorithm is better than the sequential merging algorithm. In addition, bundle adjustment merges frames better than using the linear method exclusively. Figures 6 and 7 demonstrates how the proposed method achieve more precise camera estimation and reconstruction.

In Table 1, the numbers of the selected key-frames in image sequences are 12–24. In particular, the frame number of the "Desk" sequence is 407 and 13 key-frames are selected. In the proposed LMedS-based method, because the absolute quadric is obtained from two images, a minimal subset of size (s) is 2 and the number of the possible every pair (combination) is ${}_{13}C_2 = 78$. It is often computation-

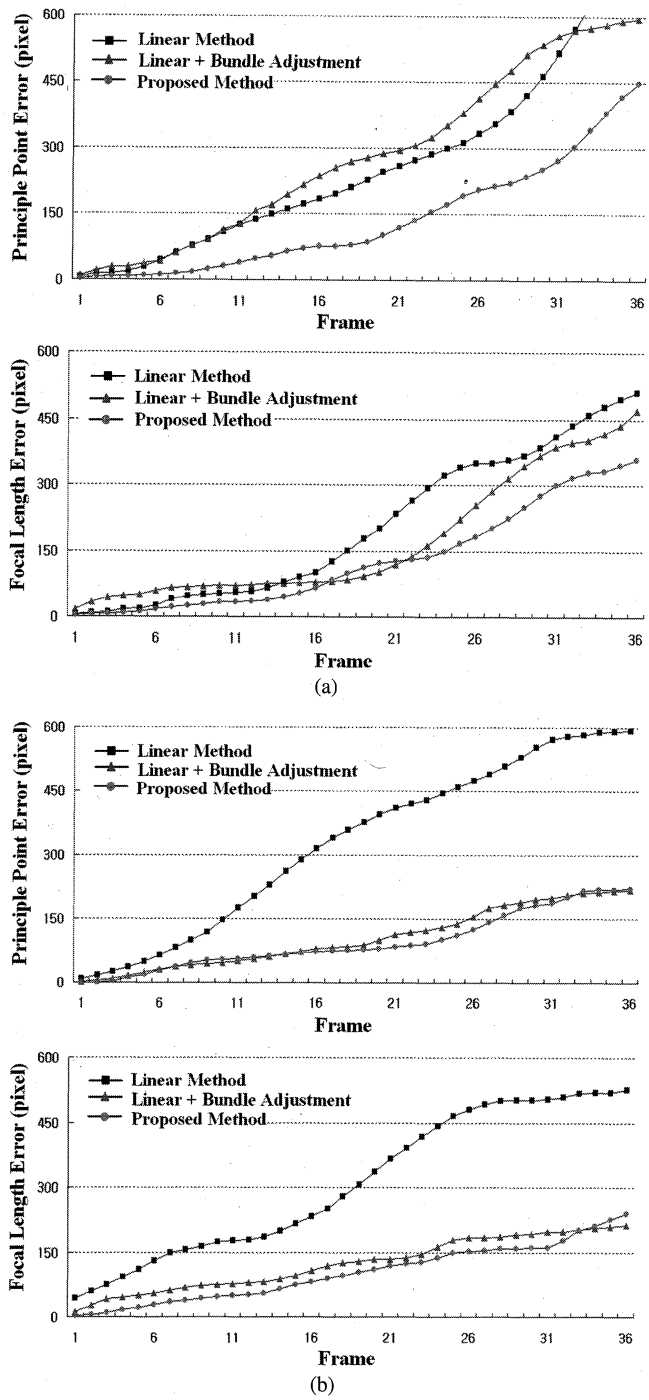


Fig. 6 Intrinsic parameter accumulation error graph, (a) on sequential merging-based reconstruction, (b) on hierarchical merging-based reconstruction.

ally infeasible and unnecessary to try every possible sample. Therefore, we should adjust a sampling rate suitable to remove the outlier effects by using the above Eq. (11).

In the simulation results, we have checked that most of the selected key-frames were generally inlier sets. However, in order to cope with the worst case: half of the key-frames are outliers ($\varepsilon = 0.5$), we determine the sampling rate as 17. If more than half the data is outlying, the LMedS-based

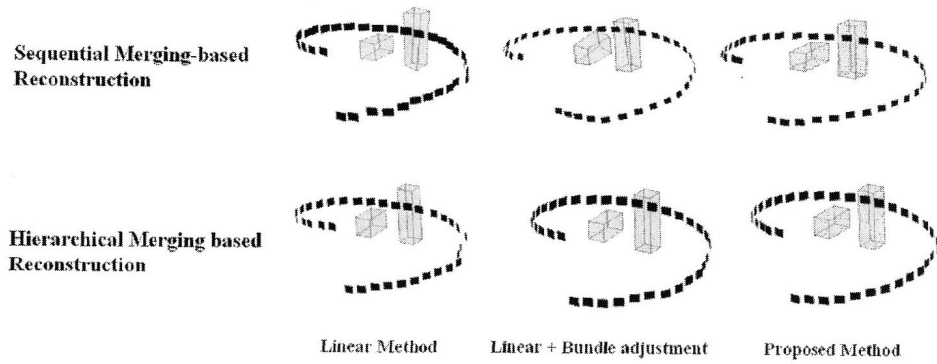
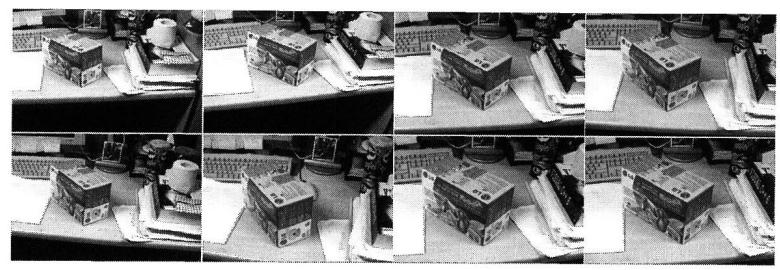
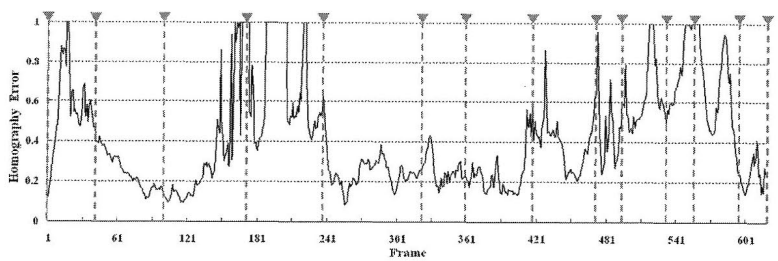


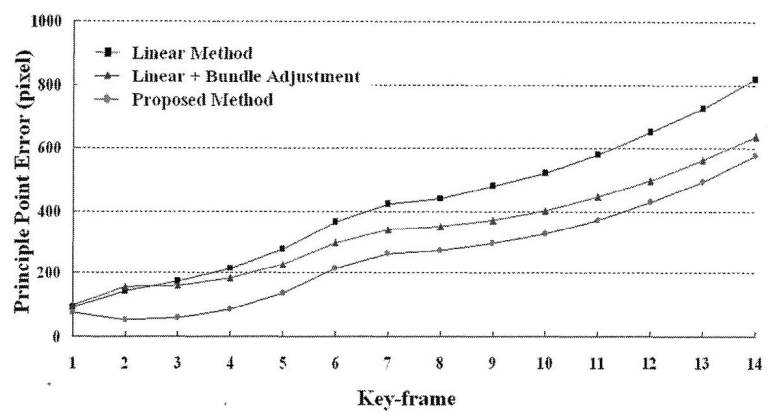
Fig. 7 Recovered camera motion and scene geometry.



(a)



(b)



(c)

Fig. 8 Simulation results on Box sequence, (a) Box sequence (1, 100, 200, 300, 400, 500, 600, and 621 frame, from upper left to lower right), (b) key-frames on the homography error, (c) principal point accumulation error of key-frames on sequential merging-based.

method fails since the median distance will be to an outlier [12]. The proposed method takes no account of this worst case, where most of the key-frames are outliers in the absolute quadric estimation, and an additional solution would be needed.

Figures 8 (a) and (b) show 8 frames in the "Box" sequence and 14 key-frames selected by the proposed method, respectively. (c) represents the results using the linear method, bundle adjustment and the proposed method, on the sequential merging. The results are obtained by accu-

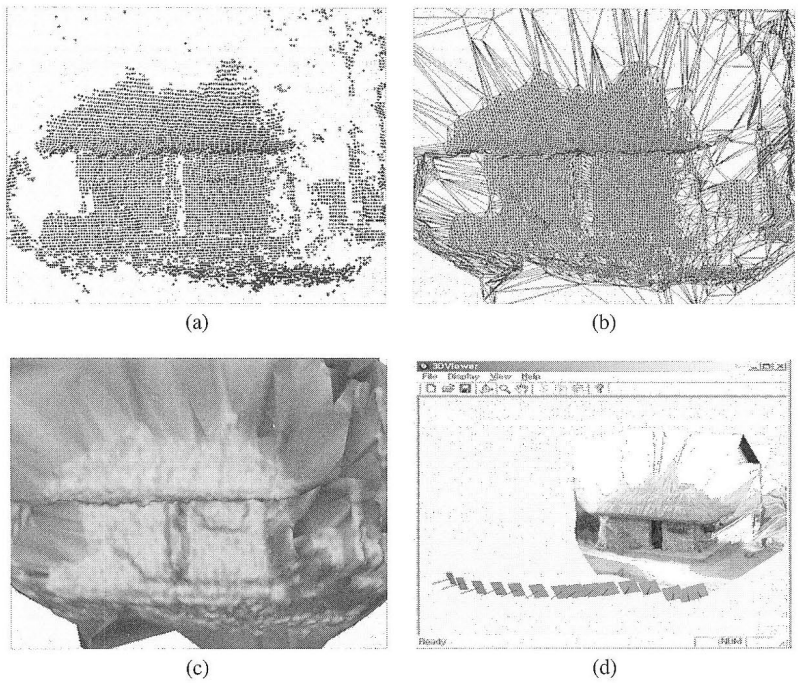


Fig. 9 3D reconstruction from Cottage sequence, (a) 3D ponts (b) triangle mesh, (c) shaded model, (d) recovered camera motion and textured model.

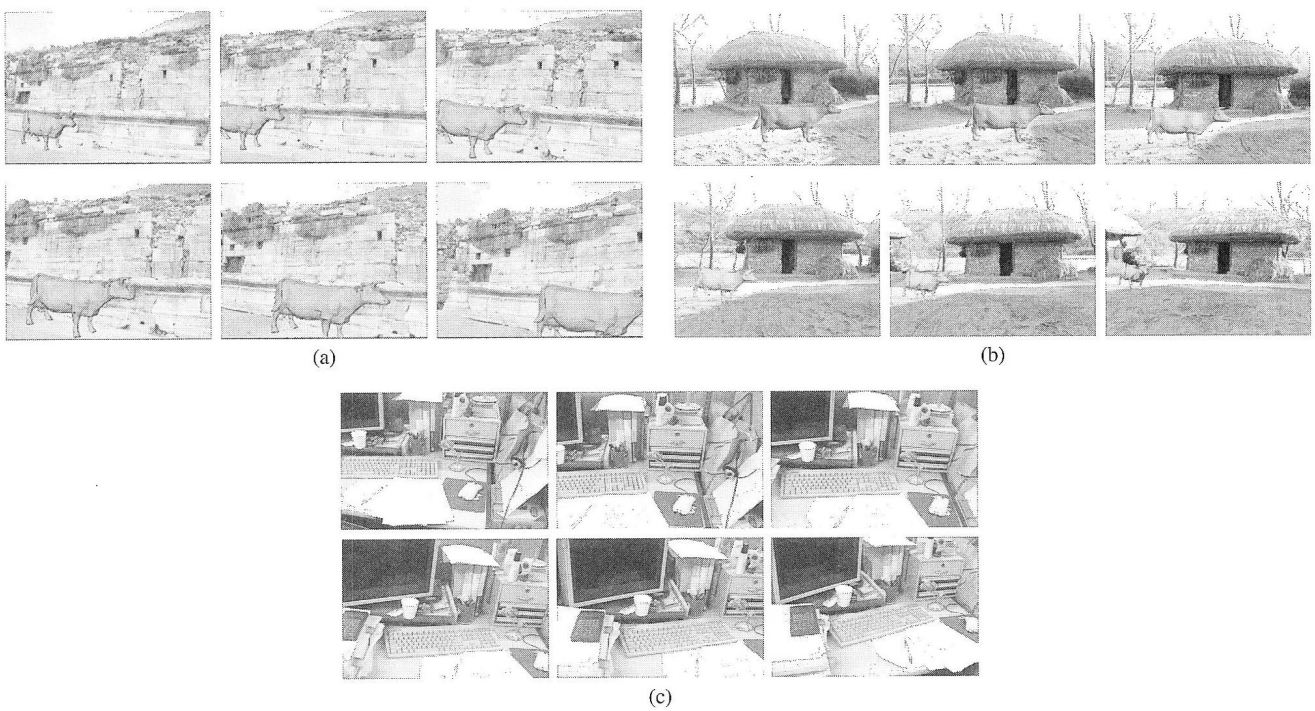


Fig. 10 Augumented image sequences, (a) Fountain, (b) Cottage, (c) Desk_2.

mulating the distance error of the actual principal point of the camera and the estimated point.

Since we have achieved calibration between successive image pairs, we can exploit the epipolar constraint that restricts the correspondence search to an 1-D range. To establish a correspondence, rectification is performed, which enables the epipolar lines to coincide with the image scan lines, followed by dense stereo matching [16]–[18]. Figure 9 shows the reconstructed 3D points with a 5×5 matching block, the triangle mesh, the shaded model, and the recovered camera trajectory and the textured model from the "Cottage" sequence, respectively. The use of the hierarchi-

cal matching on multiple views may alleviate the effects caused by occlusion, and more consideration of advanced dense matching measures can achieve better performance. The virtual object is integrated with the real scene by using the recovered camera motion and the scene geometry. Figure 10 shows augmented image sequences of "Fountain", "Cottage", and "Desk_2".

6. Conclusion

The authors propose a novel measure for key-frame selection and LMedS-based absolute quadric estimation for camera motion and scene structure recovery. By considering the ratio of the number of corresponding points and feature points, the homography error and the distribution of corresponding points, a sparse but sufficient set of views for 3D estimation can be selected with computational efficiency. In addition, the LMedS-based method enables frames with large errors among the key-frames to be rejected efficiently. Simulation results demonstrated that our algorithm achieved more precise camera and scene reconstruction of image sequences without non-linear optimization.

Acknowledgments

This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul R&BD Program, and by the Ministry of Education and Human Resources Development (MOE) under the second stage of BK21 program.

References

[1] H.S. Sawhney, Y. Guo, J. Asmuth, and R. Kumar, "Multi-view 3D estimation and applications to match move," Proc. IEEE Multi-view Modeling & Analysis of Visual Scenes, pp.21–28, 1999.

[2] O. Faugeras, Q.T. Luong, and S. Maybank, "Camera self-calibration: theory and experiments," LNCS 588, pp.321–334, 1992.

[3] W. Triggs, "Auto-calibration and the absolute quadric," Proc. IEEE Computer Vision and Pattern Recognition, pp.609–614, 1997.

[4] M. Pollefeys and L.V. Gool, "Self-calibration from the absolute conic on the plane at infinity," LNCS 1296, pp.175–182, 1997.

[5] K. Cornelis, M. Pollefeys, M. Vergauwen, and L.V. Gool, "Augmented reality from un-calibrated video sequences," LNCS 2018, pp.144–160, 2001.

[6] S. Gibson, J. Cook, T. Howard, R. Hubbold, and D. Oram, "Accurate camera calibration for off-line, video-based augmented reality," Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality, pp.37–46, 2002.

[7] P. Sturm and B. Triggs, "A factorization based algorithm for muti-Image projective structure and motion," LNCS 1065, pp.709–720, 1996.

[8] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Motion and structure causally integrated over time," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.4, pp.523–535, 2002.

[9] A. Fitzgibbon and A. Zisserman, "Automatic camera recovery for closed or open image sequences," LNCS 1406, pp.311–326, 1998.

[10] D. Nister, "Frame decision for structure and motion," LNCS 2018, pp.17–34, 2001.

[11] Z. Zhang, R. Deriche, O. Faugeras, and Q. Loung, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Technical Report of INRIA, vol.2273, 1994.

[12] R. Hartley and A. Zisserman, Mutiple View Geometry in Computer Vision, Cambrige Univ. Press, 2000.

[13] R. Hartley, "In defense of the 8-Point algorithm," Proc. IEEE International Conference on Computer Vision, pp.1064–1070, 1995.

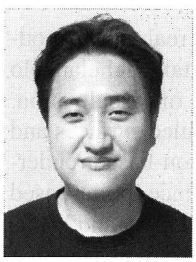
[14] J. Seo, H. Hong, C. Jho, and M. Choi, "Two quantitative measures of inlier distributions for precise fundamental matrix estimation," Pattern Recognit. Lett., vol.25, no.6, pp.733–741, 2004.

[15] P.J. Rousseeuw and A.M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, New York, 1987.

[16] Z. Zhang, "Computing rectifying homographies for stereo vision," Proc. IEEE Computer Vision and Pattern Recognition, pp.125–131, 1999.

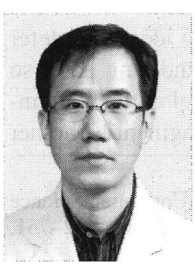
[17] C. Tsai and A.K. Katsaggelos, "Dense disparity estimation with a divide-and-conquer disparity space image technique," IEEE Trans. Multimed., vol.1, no.1, pp.18–29, 1999.

[18] S.S. Intille and A.F. Bobick, "Disparity-space images and large occlusion stereo," Proc. European Conference on Computer Vision, pp.179–186, 1994.



Yongho Hwang received his BS and MS degrees in Electronic Engineering from Myongji University, Seoul, Korea in 1996 and 1998. From 1998 to 2003, he was a junior researcher with the R&D Center of Unik C&C and Comtec System. Starting 2003, he has been a PhD student of the Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University. His research interests include computer vision and computer graphics.

Jungkak Seo received his BS degree from the Department of Architectural Engineering, DaeJeon University and his MS degree from the Department of Image Engineering, Chung-Ang University, Seoul, Korea, in 2001 and 2003, respectively. His research interests include computer vision and multi-media systems.



Hyunki Hong received his BS, MS, and PhD degrees in Electronic Engineering from Chung-Ang University, Seoul, Korea in 1993, 1995, and 1998, respectively. From 1999 to 2000 he worked at Chung-Ang Univ. as a research professor in the Information and Communication Research Institute. Since 2000, he has been a Professor of the Department of Image Engineering of the Graduate School of Advanced Imaging Science, Multimedia & Film in Chung-Ang University. His research interests include computer vision and computer graphics.