3D Reconstruction and Self-calibration based on Binocular Stereo Vision

Rongrong Hou[1] and KyungSeok Jeong[2*]

[1]**Dept. of Mechanical Engineering, Graduate School, Korea University of Tech. and Edu.**
[2]**School of Mechanical Engineering, Korea University of Tech. and Edu.**

스테레오 영상을 이용한 자기보정 및 3차원 형상 구현

후영영[1], 정경석[2*]

[1]한국기술교육대학교 대학원 기계공학부, [2]한국기술교육대학교 기계공학부

Abstract   A 3D reconstruction technique from stereo images that requires minimal intervention from the user has been developed. The reconstruction problem consists of three steps of estimating specific geometry groups. The first step is estimating the epipolar geometry that exists between the stereo image pairs which includes feature matching in both images. The second is estimating the affine geometry, a process to find a special plane in the projective space by means of vanishing points. The third step, which includes camera self-calibration, is obtaining a metric geometry from which a 3D model of the scene could be obtained. The major advantage of this method is that the stereo images do not need to be calibrated for reconstruction. The results of camera calibration and reconstruction have shown the possibility of obtaining a 3D model directly from features in the images.

요 약   스테레오 영상으로부터 3차원 형상을 구현함에 있어 사용자의 개입을 최소로 필요로 하는 기법을 개발하였다. 형상구현은 특정 기하학 그룹을 평가하는 3단계로 이루어진다. 1단계는 영상에 존재하는 epipolar 기하 평가로 각 영상에서의 특정점들을 일치시킨다. 2단계는 소실점 방법을 이용하여 투영공간에서 특정평면을 찾는 affine 기하 평가이다. 3단계에서는 카메라의 자기보정을 포함하며 3차원 모델이 얻어질 수 있는 계량 기하 변수를 구한다. 이 방법의 장점은 형상구현을 위해 스테레오 영상을 보정할 필요가 없는 것으로, 그 구현가능성을 실증하였다.

Key Words : camera self-calibration, 3D reconstruction, stereo image, epipolar geometry, affine geometry

1. Introduction

The need for 3D reconstruction of structure and motion from un-calibrated 2D images has been continuously increased in many areas such as computer vision, robot, architecture, and medicine. Significant progress has been made during the past two decades[1-3]. However, most of the efforts have been focused on the development of high accuracy and complete modeling of complex scenes based on matching primitive features, such as points and lines regardless of their relationships or context in the scene.

In recent years, many researchers have focused on the use of geometric constraints arising from the scenes to optimize the reconstruction.[4-7]. Such constraints include parallelism, orthogonality, coplanarity and other special inter-relationship of features. Werner and Zisserman developed the idea of automatically fitting a plane and modeling the perturbation surfaces of architectures[8]. The model is parameterized by combining disparity and gradient extrema and is trained from examples. This

method is based on a relatively good projective or metric reconstruction of points and lines, which may not be available in some situations. In the process of developing the technique, the need to reconstruct 3D line segments has been arisen [9-10].

Xu et al. proposed a linear algorithm for camera calibration and reconstruction from two views given homography of two planes in space[11]. The method calibrates two one-parameter cameras under the assumption of zero-skew, unit aspect ratio and the principal point being at the image center. However, this method requires a lot of human interactions.

In this research, an automatic 3D reconstruction technique from two stereo images of an object has been developed.

2. Theory

3D model reconstruction procedure from a pair of stereo images consists of following steps and the process is illustrated in Fig.1
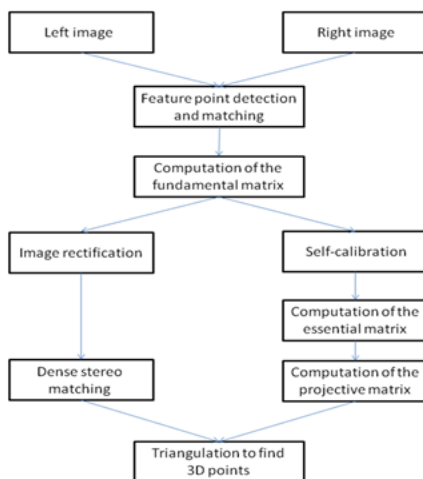
[Fig. 1] Outline of the reconstruction method

1. Features are detected in each image independently and a set of initial matching of features is calculated.
2. A fundamental matrix is calculated using the set of initial matches. False matches are discarded and the fundamental matrix is refined.
3. Projective camera matrix is established from the fundamental matrix, and upgraded to an affine camera matrix using the plane at infinity.
4. Triangulation method is used to obtain a full 3D reconstruction model with the use of metric camera matrix[12].
5. If necessary, dense stereo matching technique is applied to obtain a 3D model to be reconstructed.[13]

2.1 Feature Detection and Matching

Image matching is a fundamental problem in computer vision which includes object or scene recognition, obtaining 3D structure from multiple images, stereo correspondence, and motion tracking. In this study SIFT (Scale Invariant Feature Transform) method which transforms image data into scale-invariant coordinates relative to local features has been used. A typical image of size 500×500 pixels can generate up to about 2000 stable features depending on both image content and choices for various parameters.

Sets of image features are generated in the following computing stages.

1. Scale-space extreme detection: All scales and image locations are searched. This process can be efficiently implemented by using a difference-of-Gaussian function to identify potential interest points which are invariant to scale and orientation.
2. Key point localization: At each candidate location, a detailed model is fit to determine the location and scale. Key points are selected based on the measures of their stability.
3. Orientation assignment: One or more orientations is assigned to each key point location based on local image gradient. All operations in the next steps are performed with the image data transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.
4. Key point descriptor: Local image gradients are measured at the selected scale in the region around each key point. These are transformed into a representation which allows significant levels of local shape distortion and change in illumination.

2.2. Camera Self-calibration

2.2.1 Simplified Kruppa Equations

Simpler type variants of the Kruppa equations are applied[14]. These equations help to reduce the number of equations of six in the original formulation. And it becomes easier to select the equations to employ for self-calibration. It helps to avoid employing the epipole e' which is difficult to estimate accurately in the presence of noise and degenerate motions. And the Singular Value Decomposition (SVD) of the matrix F is employed[14]:

$$F = UDV^T \tag{2.1}$$

Recalling that F is of rank two, the diagonal matrix D has the following form:

$$D = \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.2}$$

where r and s are the eigenvalues of the matrix FF^T, whereas U and V are two orthogonal matrices.

As D is a diagonal matrix having the last element of zero, the following direct solution for e' is obtained:

$$e' = \delta U m, \quad \delta \neq 0 \tag{2.3}$$

with $m = [0, 0, 1]^T$. Then, the matrix $[e']_\times$ is equal to

$$[e']_\times = \mu U M U^T \tag{2.4}$$

Substitution of Eq. (2.4), yields a new expression for the Kruppa equations:

$$FKF^T = v U M U^T K' U M^T U^T, v \neq 0 \tag{2.5}$$

Since U is an orthogonal matrix, left and right multiplication of Eq.(2.5) by U and U^T respectively, yields the following notably simple expression for the Kruppa equations:

$$DV^T KVD^T = v M U^T K' U M^T \tag{2.6}$$

Because of the simple form of the matrices D and M, Eq.(2.6) corresponds to three linearly dependent equations. Practically denoting the column vectors of U by u_1, u_2, u_3 and by v_1, v_2, v_3, the matrix equation (2.6) is equivalent to

$$DV^T KVD^T = \begin{bmatrix} r^2 v_1^T K v_1 & rs v_1^T K v_2 & 0 \\ rs v_2^T K v_1 & s^2 v_2^T K v_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M^T K' U M^T = \begin{bmatrix} u_2^T K' u_2 & -u_2^T K' u_1 & 0 \\ -u_1^T K' u_2 & u_1^T K' u_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The above expressions finally yield the following three linearly dependent equations for the matrices K and K':

$$\frac{r^2 v_1^T K v_1}{u_2^T K' u_2} = \frac{rs v_1^T K v_2}{-u_2^T K' u_1} = \frac{s^2 v_2^T K v_2}{u_1^T K' u_1} \tag{2.7}$$

Only two out of these three equations are linearly independent. They are the simplified Kruppa equations derived in a particularly straightforward manner. Moreover, the use of SVD should be taken into account to deduce three equations out of six in the original formulation. It should be noted that the simplified Kruppa equations are not symmetric with respect to the pair of images used. Since the fundamental matrix defined by reversing the role of the images in a pair is equal to $F^T = VDU^T$, the analogous of Eq.(2.6) becomes $DV^T KVD^T = v M U^T K' U M^T$, which is different from Eq.(2.6).

2.2.2 Intrinsic Parameter with the use of the Simplified Kruppa Equation

Once the fundamental matrix is recovered the camera intrinsic matrix can be recovered. For the time being, only the case of self-calibration for unknown focal length is considered to be implemented. This is based on a general assumption that, for modern well engineered cameras, the principal point (p_x, p_y) is at the centre of the image and the aspect ratio is one. It is also assumed that the intrinsic matrixes of the images (no zoom) are the same, so

$$A = A^{'} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } K = K^{'} = \begin{bmatrix} f^2 & 0 & 0 \\ 0 & f^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2.3 Extrinsic Parameters

The fundamental matrix F describes the epipolar geometry between the pair of views considered. It is the equivalent to the essential matrix $E = [t]_{\times} R$ for the uncalibrated case, as dictated by

$$F = A^{'-T} E R A^{-1} \tag{2.8}$$

Due to the above relation, E can be written as a function of F as follows:

$$E = A^{'T} F A \tag{2.9}$$

The essential matrix, has only five degrees of freedom: both the rotation matrix R and the translation t have three degrees of freedom each, still there is an overall scale ambiguity like the fundamental matrix, the essential matrix is a homogeneous quantity[15].

The symmetric matrix EE^T is independent of the rotation R since

$$EE^T = [t]_{\times} RR^{T}([t]_{\times})^{T} = [t]_{\times}([t]_{\times})^{T} \tag{2.10}$$

$$FKF^T = A^{'-T}[t]_{\times}([t]_{\times})^{T}A^{'-T} \tag{2.11}$$

where K is the symmetric matrix AA^{T}. This equation will be employed in deriving the Kruppa equations algebraically.

Comparing this with the relation $m^{'T} F m = 0$ for the fundamental matrix gives the relationship between the fundamental and essential matrices as

$$E = K^{'T} F K \tag{2.12}$$

The extrinsic parameters can be extracted when the intrinsic parameters are determined using the non-linear algorithms.

3. Experimental Results and Analysis

3.1 3D Reconstruction

If the two points in two images satisfy the epipolar constraint, $m^{'T} F m = 0$, and camera matrices P and $P^{'}$ are known, the epipolar constraint ensures the existence of the epipolar plane which contains the backward projected rays from m and $m^{'}$. When the two rays share a common plane, they intersect at a point X. The projection of this point X using camera matrices P and $P^{'}$ gives m and $m^{'}$. Then the triangulation method is applied to estimate 3D points from a pair of points in two images.

If camera matrices (P and $P^{'}$) and 2D feature points (m and $m^{'}$) are given, 3D points can be calculated using non iterative Linear triangulation method. The projection equations $m = PM$ and $m^{'} = P^{'}M$ can be expressed as a linear system of equation $AM = 0$. Then the homogeneous scale factor can be eliminated by a cross product and each image point provides three equations, two of which are linearly independent. For instance $m \times (PM) = 0$ yields

$$u(P^{3T}M) - (P^{1T}M) = 0$$
$$v(P^{3T}M) - (P^{2T}M) = 0$$
$$u(P^{2T}M) - v(P^{1T}M) = 0$$

where P^{iT} are the rows of camera matrix P [12]. This can also be written as $[m] \times (PM) = 0$. Now considering both projections, a system of the form $AX = 0$ can be composed, where

$$A = \begin{bmatrix} [m] \times P \\ [m^{'}] \times P^{'} \end{bmatrix}$$

3.2 Reconstruction Results

Two stereo images of an object can be acquired by taking images either using two cameras at the same time or one camera at a sequential time. In this study, pair of images has been acquired with the sequential use one camera. The resolution of the camera used for the measurement was 1.2M pixels (3,872 × 2,592).

Case 1: The technique developed has been applied for a cuboid. The pair of images used for 3D reconstruction

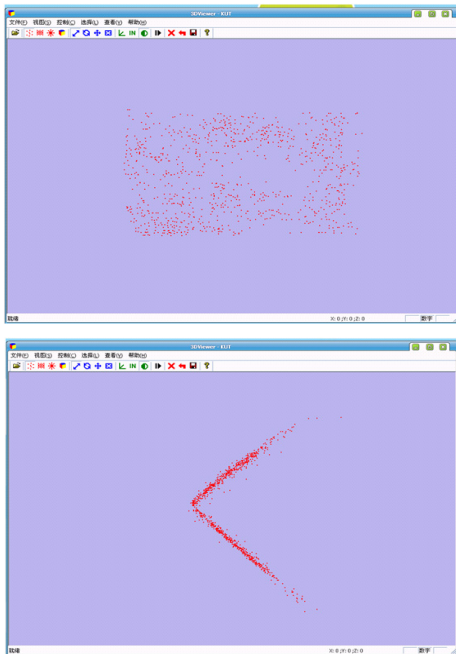are shown in Fig. 2. All matching points detected are shown in Fig. 3.
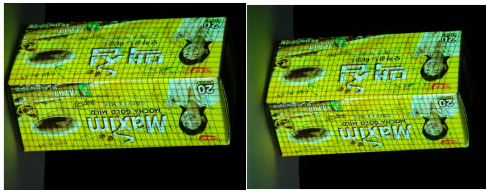
[Fig. 2] Stereo Image Pair (640 × 480 pixels)
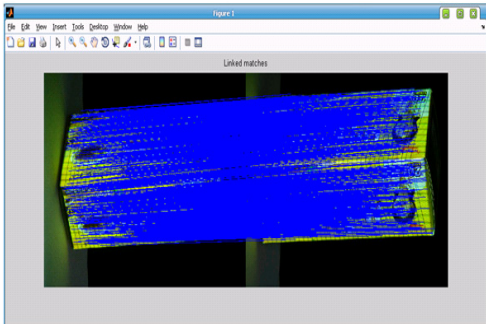
[Fig. 3] Corresponding Points of Stereo image

Reconstructed 3D points seen from the front and the top are shown in Fig. 4.
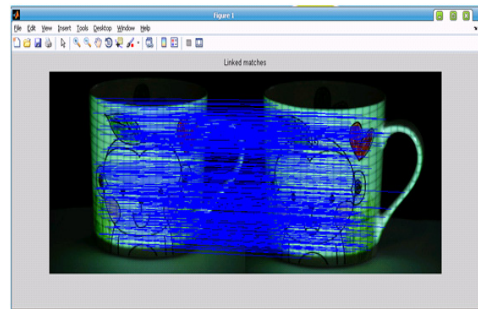

[Fig. 4] Reconstructed 3D points of the cuboid : Front and top view

Case 2: Another trial has been made for an object with curved surface. The pair of images used is shown in Fig. 5. And the matching points detected from the stereo images of the cup are shown in Fig. 6.
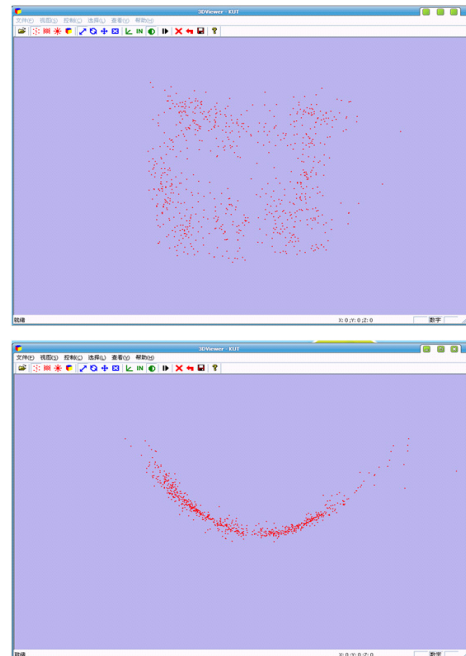
Reconstructed 3D points of the cup are shown in Fig. 7.

[Fig. 5] Stereo Image Pair (Left and Right image, 640 × 480 pixels)

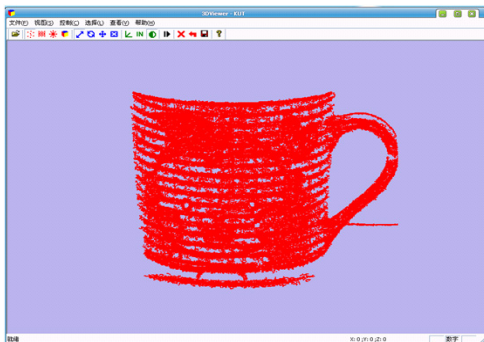

[Fig. 6] Corresponding points of stereo images

[Fig. 7] Reconstructed 3D points of the cup : Front and top view

3.3 3D Textured Model

As the matching algorithm for un-calibrated stereo images gives only a few points matched. a dense stereo matching method has been applied to increase the number of matching points. For dense stereo matching[1,13], stereo images need to be rectified for the search space to be one dimensional. After the image rectification, matching points in the second image can be obtained in almost all corresponding pixels in the first image.

The 3D textured model is obtained from each pixel of the disparity map. The disparity index works as an index for the left image and the value of disparity as horizontal offset of the index in the right. With the aid of the rectified camera projection matrices and the triangulation method, each point is reconstructed and assigned with the average pixel value from both images. The reconstructed 3D points are shown in Fig. 8. Points linked based on Delaunay triangulation are shown in Fig. 9.

knowledge available on the parameters of the stereo system. If the intrinsic parameters and extrinsic parameters are available, the metric reconstruction can be made with a scale factor. The scale factor can be determined with the known distance between two points of the observed image.

The metric structure preserves not only parallelism but also angles and length ratios. Thus the ratio of a certain length on real object and its corresponding image can give a scale factor, which enables the estimation of the dimension of the real object.

To assess the accuracy of the technique, six sets of corresponding lengths of the cuboid are both measured on the real object and calculated in its two images. The measured edge lines are shown in Fig. 10 and the results are given in Table 1. The comparison gives that the average relative error is less than 2.5% with the maximum of 5.05%.

[Fig. 8] Reconstructed 3D points of a cup


[Fig. 10] Edges of a cuboid measured

[Table 1] Comparison of reconstructed distances and real distance

	AB	CD	EF	AE	BF	AC	BD
Real distances /mm	174	174	174	80	80	69	69
Reconstructed distances/ mm	174	173.4152	170.8277	76.7112	79.4734	65.5131	70.3524
Absolute error/mm	0	0.5848	3.1723	3.2888	0.5266	3.4869	1.3524
Relative error/%	0	0.336	1.823	4.111	0.6582	5.05	1.96

[Fig. 9] Reconstructed 3D texture model

3.4 Scale Factor

Successful 3D reconstruction depends on the priori

Results of measurement using this method are shown in Figure 11(a) and Fig. 11(b), and it can be observed that

outlines of the paper box in different profiles are preserved after the 3D reconstruction.

Fig. 12(a) and Fig. 12(b) are the results of 3D reconstruction of a cup and sectional profiles are compared with those of actual profiles. Only small deviation in the profiles can be observed in the comparison. And this method proved to be applied in the measurements of length, width and depth of real objectives.
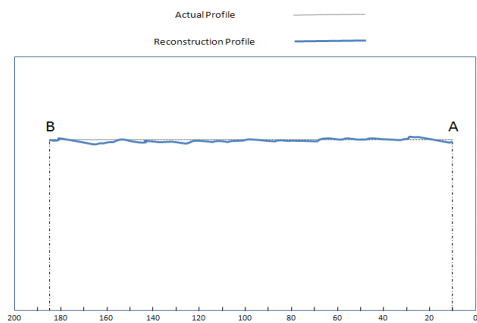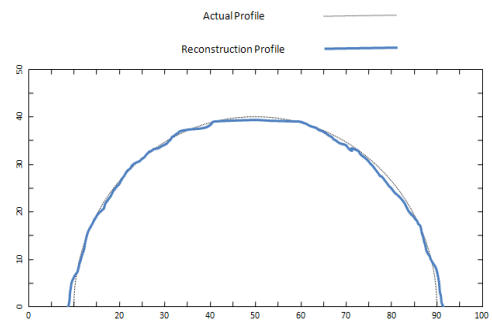

[Fig. 11(a)] A-B outline of the box measured
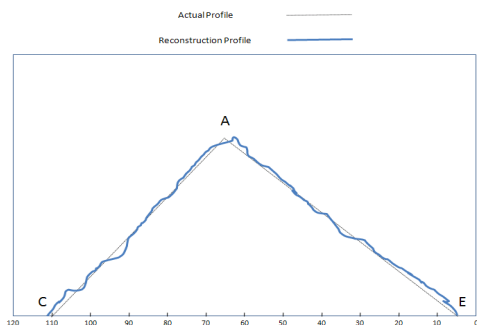
[Fig. 11(b)] C-A-E outline of the box measured
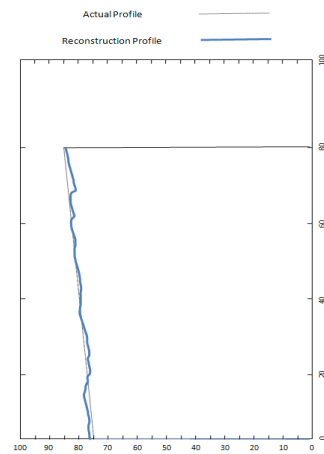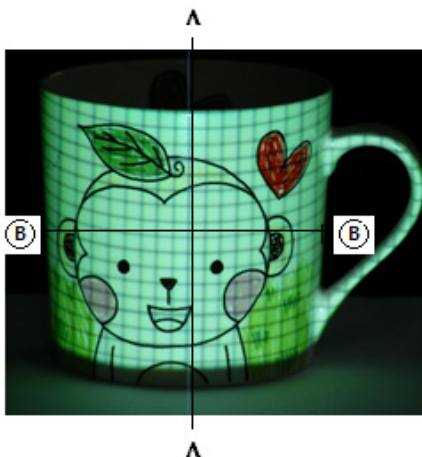
(a) A-A section

(b) B-B Section

[Fig. 12] Cross sectional profile of the cup measured

4. Conclusion

1. A 3D reconstruction technique from two stereo images, which requires minimal intervention of the user, has been developed.
2. Simplified Kruppa equations enable the selection of equations for self-calibration and helps to avoid employing epipole in the presence of noise.
3. Test results show that this technique can be applied with acceptable level of errors.

References

[1] B. Han, C. Paulson, D. Wu, "3D dense reconstruction from 2D video sequence via 3D geometric segmentation", J. of Visual Communication and Image

Representation, vol. 22, Issue 5, pp. 421-431, 2011.

[2] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, M. A.G. Ballester, "A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images, Medical Image Analysis, vol. 13, Issue 6, pp. 883-899, 2009.

[3] R. I. Hartley, and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.

[4] G. Trombettoni, M. Wilczkowiak, "Gpdof - a Fast Algorithm to Decompose under-constrained Geometric Constraint Systems: Application to 3d Modeling", Int. J. Comput. Geometry Appl., vol. 16(5-6), pp. 479-512, 2006.

[5] R. Szeliski and P. H. S. Torr, "Geometrically Constrained Structure from Motion: Points on Planes" Reinhard Koch and Luc Van Gool Eds., SMILE '98, LNCS 1506, pp. 171-186, 1998.

[6] W. Zhang and J. Kosecka, "Extraction, matching and pose recovery based on dominant rectangular structures", High Level Knowledge in Vision Workshop, International Conference on Computer Vision, Nice France, 2003.

[7] R. Hartley and J. L. Mundy, "The relationship between photogrammetry and computer vision", Proc. SPIE Vol. 1944, pp. 92-105, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision, Eamon B. Barrett; David M. McKeown; Eds. 1993.

[8] T. Werner and A. Zisserman, "New techniques for automated architecture reconstruction from photographs", Proc. 7th Eur. Conf. Comput. Vision, vol. 2, pp. 541 – .555. 2002a.

[9] C. Baillard and A. Zisserman, A, "Automatic Reconstruction of Piecewise Planar Models from Multiple Views", Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 1999.

[10] P. Gargallo, E. Prados, P. Sturm, "Minimizing the Reprojection Error in Surface Reconstruction from Images", IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007.

[11] G. Xu, J. Terai and H. Shu, "A linear algorithm for Camera Self-Calibration, Motion and Structure Recovery for Multi-Planar Scenes from Two Perspective Images, http://research.microsoft.com/apps/pubs/default.aspx?id=68822, 2000.

[12] R. Hartley, "Euclidean Reconstruction from Uncalibrated Views", in J. L. Mundy, A. Zisserman, and D. Forsyth Eds., 'Applications of Invariance in Computer Vision', volume 825 of Lecture Notes in Computer Science, pp. 237-256, Springer-Verlag, 2004.

[13] H. Sekkati and A. Mitiche, "A variational method for the recovery of dense 3D structure from motion", Robotics and Autonomous Systems, vol. 55, Issue 7, pp. 597-607, 2007.

[14] M. Lourkis and R. Deriche, "Camera Self-Calibration Using the Kruppa Equations and the SVD of the Fundamental Matrix: The Case of Varying Intrinsic Parameters", INRIA report No. 3911, 2000.

[15] H. P. Trivedi, "Can multiple views make up for lack of camera registration", Image and Vision Computing, 6(1): 29-32, 1988.

Kyung-Seok Jeong   [Regular member]

- Feb. 1989 : Seoul Natl. Univ., Aeronautical Engrg., PhD
- Feb. 1989 ~ Apr. 1990 : Daewoo Motors Co., Senior Researcher
- Apr. 1990 ~ Feb. 1992 : Imperial College of Science & Technology (UK), Researcher
- Sep. 1992 ~ current : Korea Univ. of Tech. & Edu., Dept. of Mech. Engrg., Professor

<Research Interests>
3D Images, Thermo-fluids Phenomena, Laser Diagnostics

Rongrong Hou   [Regular member]

- Sep. 2002 : Yan Shan Univ. (CHN)., Photoelectron Department, Bachelor
- Sep. 2008 : Korea Univ. of Tech. & Edu., Dept. of Mech. Engrg., MS
- Sep. 2010 ~ current : JT Company, Vision Development Team, Engineer

<Research Interests>
3D Reconstruction, Image Processing