Monocular Three-dimensional Self-calibrating Surface Reconstruction using Digital Photogrammetry

Tho, N.T.[1], Suada, M.G.[1], Dirgantara, T.[1], and Putra, I.S.[1*]

[1]*Lightweight Structure Research Group*
Bandung Institute of Technology, Bandung, Indonesia, 40132

Photogrammetry, which has significantly matured throughout the years, is an optical technique to accurately measure the depth of spatial points from their images. In this paper, a self-calibrating measurement technique based on photogrammery is presented. A digital camera with zoom lenses is used to capture images of the investigated object. The image sequence is then transferred to a Matlab® code to detect the interest points, then to compute the 3-D coordinates of those interest points and the camera poses by matrix factorization operation. Neither calibration object nor *a priori* information of the camera is needed. The proposed 3-D reconstruction technique is proved able to measure the 3-D shapes of various surfaces with the error of 1 part in 1000 of their sizes.

Keywords :

3-D optical method, self-calibration, monocular camera sequence, photogrammetry, matrix factorization.

1. Introduction

Since the years 1960s, many optical methods have been developed to solve the problems of measuring. Photogrammetry, an optical measurement method arising from geodesy, has been used in 3-D measurement for the early purpose of exactly determining any coordinate in space. It is currently well known that photogrammetry offers a comfortable way to solve the problem of shape measurement by combining with other advanced techniques such as computer vision or fringe projection. All of these standard methods required the images of investigated objects captured by cameras at different positions, and detected features from those images will be used to recover the spatial information of the objects.

These standard optical methods, however, usually measure the coordinates indirectly, when the parameters which describe the geometrical conditions of the cameras are known through camera calibration before the actual 3-D measurement. Such calibration work requires perfect standards of the calibration procedure and of the calibration objects. Furthermore, the calibration becomes more difficult with increasing object volume, and the measuring system is expected to be stable in terms of time because the calibration is done before the real measurement. Moreover, the calibration itself is sophisticated and time consuming. This state of affairs causes optical measuring systems to be complicated to handle and sensitive to environmental changes due to disturbances during the time between the calibration and the measurement. Therefore, a self-calibrating reconstruction algorithm without camera calibration is beneficial in such situation.

Geometric and algebraic relations among uncalibrated views (up to four) have been described in [1]. Various algorithms for scene reconstruction with both orthographic and perspective camera have been proposed. Tomasi and Kanade [2] developed a factorization method of the measurement matrix for scene reconstruction with an orthographic camera, and Sturm and Triggs [3] extended this method from affine to perspective projections without occlusions. Jacobs [4] solves reconstruction with occlusions for orthographic camera.

In this work, a self-calibrating measurement technique based on photogrammetry is developed. A digital camera with zoom lenses is used to capture images of the target object. The image sequence is transferred directly to a MATLAB code to detect the interest points, then to compute the 3-D coordinates of those interest points and the camera poses by matrix factorization operation. Neither calibration object nor *a priori* information of the camera is needed. To achieve high accuracy and reliability, several loops and iterations are generated in order to minimize the error cost function. The experimental results show that the proposed 3-D reconstruction technique is able to measure the 3-D shapes of various surfaces with the final error of about 1 part in 1000 of the

* Corresponding Author. E-mail : isp@aero.pauir.itb.ac.id
Phone : +62-22-4254016, Fax. : +62-22-2508519

object size. The main features of the proposed measuring technique are versatility, simplicity and inexpensiveness.

The following section briefly explains the background of the mathematical model used based on photogrammetry and how it can be applied to the proposed factorization technique. Section 3 presents some experimental results of surface reconstruction with proposed algorithm, and some discussions are opened in the last section.

2. Methodology

In this section, brief mathematical formulation related to this work and the flow chart of the proposed algorithm for surface reconstruction are introduced.

A computer-vision-based pinhole camera is modeled by its *optical center* **C** and its *image plane* with local coordinates (x, y). A spatial point **X** projects an image point **x** in the image plane, which should be expressed in equivalent homogeneous point **u**. Let **X** = (X, Y, Z, 1) be homogeneous coordinates of point **X** in the camera coordinate system (Figure 1), which can easily be transformed to the object coordinate system by a rotation matrix and a translation vector; and **u** = (u, v, 1) the homogeneous form of image point **x**. The mathematical relation between the 3-D point and its image can then be written in an *image equation* as

$$\lambda \mathbf{u} = \mathsf{P}\mathbf{X} \tag{1}$$

where the scalar λ is the projective depth of corresponding image point and P is a 3x4 projection matrix which projects **X** to **u**. The image equation (1) can be rewritten in the case of multiple views as

$$\lambda_j^i \mathbf{u}_j^i = \mathsf{P}^i \mathbf{X}_j \, . \tag{2}$$

in which the superscript stands for the index of image that the point $\mathbf{u}_j$ is seen, and the subscript for the index of point appeared in the image i. In this case, the image points **u** are known, and reconstruction is the work to recover P (the motion) and **X** (the structure) from the image points. Classically, P^i is determined by camera calibration before finding $\mathbf{X}_j$. However, due to some disadvantages of calibration shown in previous section, it is more beneficial to find another way to reconstruct **X** and P at the same time by matrix factorization.

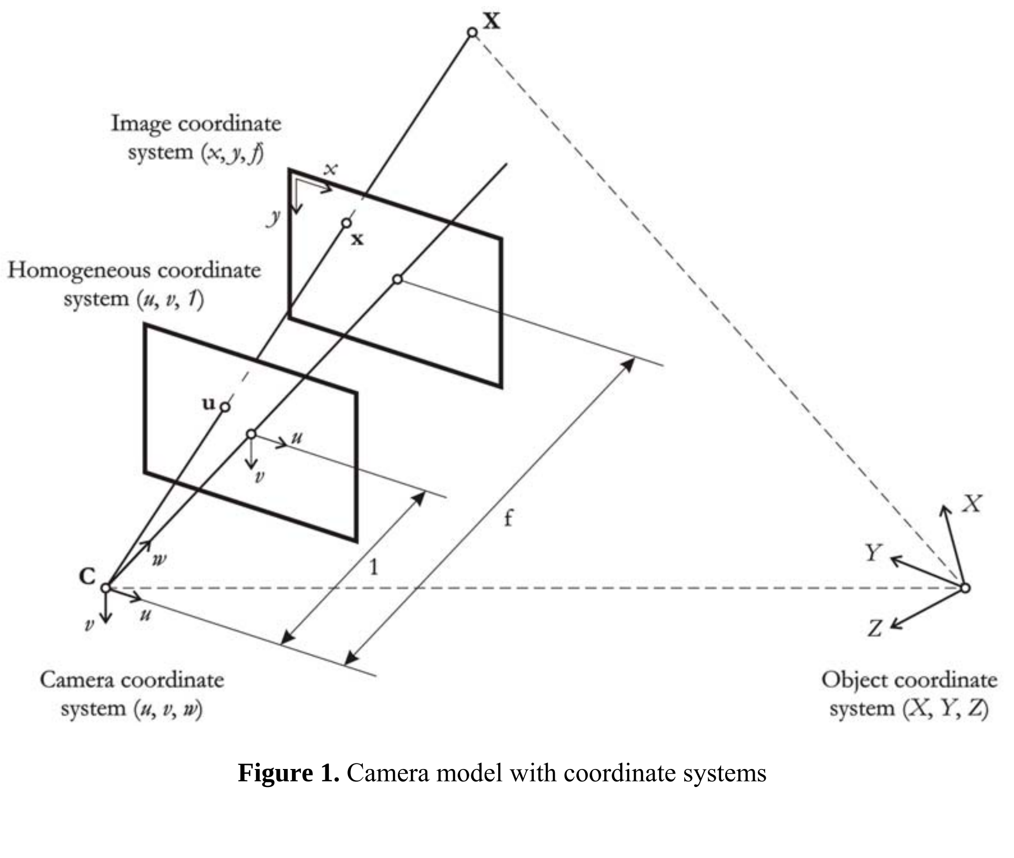


Figure 1. Camera model with coordinate systems

2.1. Measurement matrix from interest points

Suppose that n points in m images are used to reconstruct an investigated surface. Firstly, the n interest points are selected on all images. Their 2-D coordinates are transformed to homogeneous form, and then gathered together in a *measurement matrix* W so that each column of the matrix indicates a single point in every image and each row shows all points seen from one camera position:

$$\mathsf{W} = \begin{bmatrix} \mathbf{u}_1^1 & \mathbf{u}_2^1 & \dots & \mathbf{u}_n^1 \\ \mathbf{u}_1^2 & \mathbf{u}_2^2 & \dots & \mathbf{u}_n^2 \\ \vdots & & & \vdots \\ \mathbf{u}_1^m & \mathbf{u}_2^m & \dots & \mathbf{u}_n^m \end{bmatrix} \tag{3}$$

in which the superscript stands for the index of image that the point $\mathbf{u}_j$ is seen ($j = 1..n$) and the subscript for the index of point appeared in the image i ($i = 1..m$).

2.2. Projective depth estimation

It is obvious that from Eq. (1) the image points are scaled by projective depths λ which are still unknown up to now. Therefore, the depths λ should be estimated before the factorization. Sturm and Triggs algorithm [3] exploiting the epipolar geometry (Figure 2) is used as it is easy to estimate the projective depths and can be summarized in two following steps:

1. Set the depths λ_j^c for known points $\mathbf{u}_j^c$ to 1.
2. For each image $i \neq c$, compute the fundamental matrices F^{ic} (by robust RANSAC algorithm [5]), the epipoles $\mathbf{e}^{ic}$ (right-null vector of F^{ic}) and the depths λ_j^i according to

$$\lambda_j^i = \frac{\left(\mathbf{e}^{ic} \times \mathbf{u}_j^i\right)\left(\mathsf{F}^{ic}\mathbf{u}_j^c\right)}{\left\|\mathbf{e}^{ic} \times \mathbf{u}_j^i\right\|^2} \lambda_j^c \,. \tag{4}$$

Such equations (4) can be iteratively solve to give estimates for the complete set of depths for point j, starting form arbitrary initial value such as $\lambda_j^i = 1$. The iteration is stopped once the increment is smaller than a pre-defined threshold value.

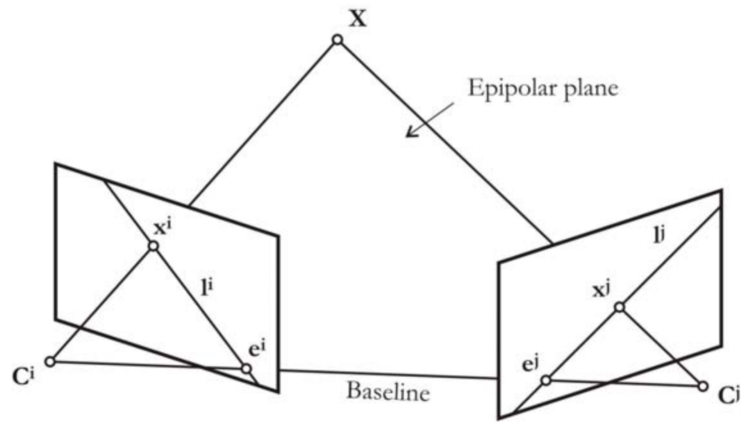


Figure 2. Epipolar geometry. The spatial point $\mathbf{X}$ is imaged in two image planes at $\mathbf{x}^i$ and $\mathbf{x}^j$. The baseline $\mathbf{C}^i\mathbf{C}^j$ intersects the image planes at the epipoles $\mathbf{e}^i$ and $\mathbf{e}^j$. Corresponding points $\mathbf{x}$ and $\mathbf{e}$ lie on a epipolar line $\mathbf{l}$. The matched pair $\mathbf{x}^i \leftrightarrow \mathbf{x}^j$ is mathematical represented though a fundamental matrix F^{ij} so that $\left(\mathbf{x}^j\right)^{\mathrm{T}} \mathsf{F}^{ij}\mathbf{x}^i = 0$.

2.3. Projective structure and motion by matrix factorization

Once the projective depths of every point are determined, the complete set of image projections can be collected in a *rescaled measurement matrix* R of size $3m\mathrm{x}n$. The image equation in case of multiple views (2) becomes

$$\mathsf{R} = \begin{bmatrix} \lambda_1^1 \mathbf{u}_1^1 & \lambda_2^1 \mathbf{u}_2^1 & \dots & \lambda_n^1 \mathbf{u}_n^1 \\ \lambda_1^2 \mathbf{u}_1^2 & \lambda_2^2 \mathbf{u}_2^2 & \dots & \lambda_n^2 \mathbf{u}_n^2 \\ \vdots & & & \vdots \\ \lambda_1^m \mathbf{u}_1^m & \lambda_1^m \mathbf{u}_2^m & \dots & \lambda_n^m \mathbf{u}_n^m \end{bmatrix} = \begin{bmatrix} \mathsf{P}^1 \\ \mathsf{P}^2 \\ \vdots \\ \mathsf{P}^m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_n \end{bmatrix}. \tag{5}$$

The idea of the proposed algorithm is revealed hereby: factorize the matrix R into projective cloud $\mathbf{X}_j$ and projective motion from P^i. According to Jacobs [4], the noise-free matrix R is of rank 4. However in presence of noise, R will only be approximately of rank 4, and it is required to improve the factorization results by balancing R. The goal is to ensure good numerical conditions of rescaling so that all rows and columns of R are of uniform magnitudes by using the following iterative scheme:

1. Rescale each column of R so that the sum of squared coordinates of the column is 1.
2. Rescale each triplet of rows of R so that the sum of squared coordinates of the triplets is 1.
3. If the entries of R change significantly and the iteration number is small enough, repeat 1 and 2.

After confirming that R is in good condition for factorization, one may decompose the matrix into P and $\mathbf{X}$ (remove indices i and j for simplicity) by *Singular Value Decomposition* (SVD) so that

$$\mathsf{R} = \mathsf{UDV} = \mathsf{U'D'V'} = \left(\mathsf{U'}\sqrt{\mathsf{D'}}\right)\left(\sqrt{\mathsf{D'}}\mathsf{V'}\right) = \bar{\mathsf{P}}\bar{\mathbf{X}} \tag{6}$$

where (U, D, V) are matrices decomposed from R by SVD definition, (U', D', V') are matrices built from the first 4 columns of (U, D, V) respectively, and ($\bar{\mathbf{X}}$,$\bar{\mathsf{P}}$) are projective cloud and projection matrices.

2.4. Auto-calibration

The factorization in Eq. (6) might be interpreted in another way by inserting any non-singular 4x4 matrix H to get another compatible structure and motion pair P and $\mathbf{X}$ as follows:

$$\mathsf{R} = \bar{\mathsf{P}}\bar{\mathbf{X}} = \left(\bar{\mathsf{P}}\mathsf{H}\right)\left(\mathsf{H}^{-1}\bar{\mathbf{X}}\right) = \mathsf{P}\mathbf{X} \tag{7}$$

In other words, the factorization recovers the projective structure and motion up to a projective transformation (or homography) H. The auto-calibration process computes such a matrix H that upgrades the projective reconstruction into the metric one, which is sometimes called Euclidean stratification. The process is based on the definition of the *Image of the Absolute Conic* (IAC) which is described in detail by Svoboda *et. al* [6].

2.5. Bundle adjustment

Bundle adjustment (BA) is the last step to refine a visual reconstruction to produce *jointly optimal* 3-D structure and viewing parameter estimates. *Optimal* means that the parameter estimates are fond by minimizing some cost function that quantified the model fitting error, and *jointly* that the solution is simultaneously optimal with respect to both structure and camera variations [7]. BA is really a large sparse geometric parameter estimation problem, the parameters being the combined 3-D coordinates and camera poses. BA minimizes the reprojection errors in terms of all 3-D points and camera parameters, specially:

$$\min_{\mathsf{P}^i, \mathbf{X}_j} \sum_{i=1}^{m} \sum_{j=1}^{n} d\left(\mathbf{Q}\left(\mathsf{P}^i, \mathbf{X}_j\right), \mathbf{x}_j^i\right)^2 \tag{8}$$

where $\mathbf{Q}\left(\mathsf{P}^i, \mathbf{X}_j\right)$ is the predicted projection of point $\mathbf{X}_j$ on image i by P^i, and $d(\mathbf{x}, \mathbf{y})$ denotes the Euclidean distance between the inhomogeneous image points represented by **x** and **y**. The expression (8) can be treated as a non-linear minimization based on Levenberg-Marquardt algorithm by solving the normal equation

$$\mathbf{J}^{\mathrm{T}}\mathbf{J}\delta = -\mathbf{J}^{\mathrm{T}}\varepsilon \tag{9}$$

in which **J** is the Jacobian of the error function, δ is the update to the parameter vector and ε is the difference between the real and the initially estimated values [8]. Note that if κ and ν are respectively the dimensions of each P^i and $\mathbf{X}_j$, the total number of minimization parameters in (8) equals $m\kappa+n\nu$ and is therefore large even for moderately sized BA problems.

The self-calibrating algorithm of surface reconstruction using digital photogrammetry is explicitly depicted in the flow chart below (Figure 3).

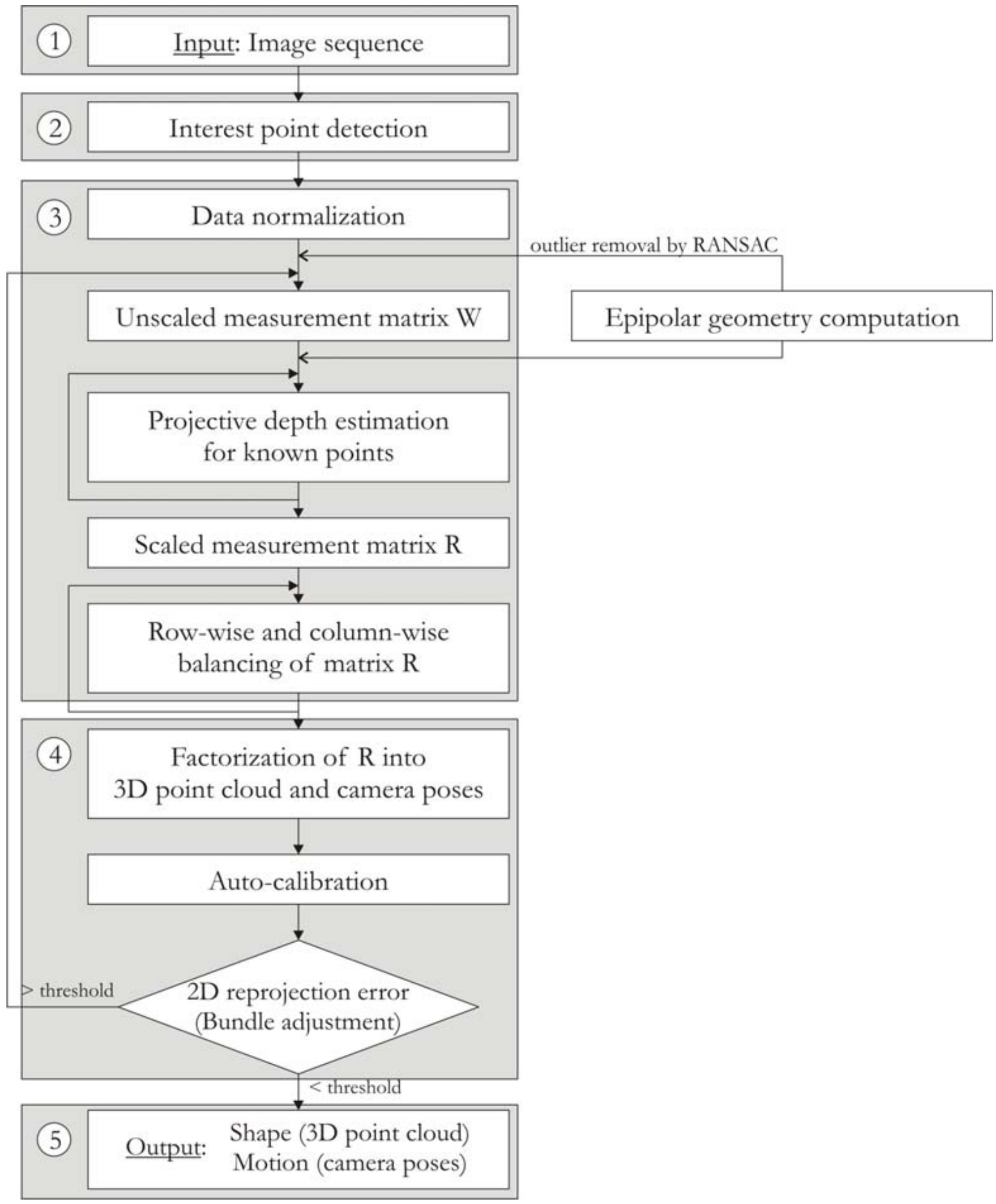


Figure 3. The proposed 3-D reconstruction algorithm chart.
Block 1: Capturing object image. Block 2: Interest point selection. Block 3: Processing for rescaled measurement matrix. Block 4: Processing for 3-D points and camera poses. Block 5: Post-processing to display the results.

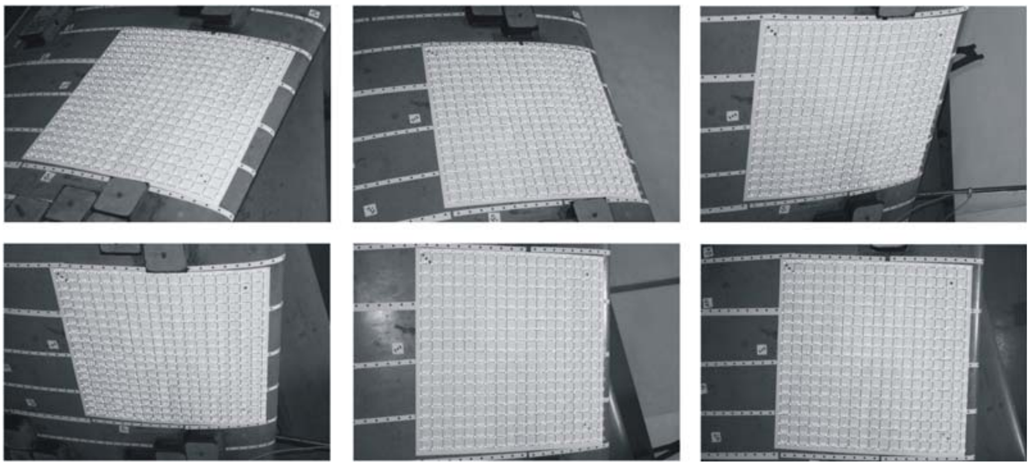
3. Experimental results

The experiment for the 3-D surface measurement was performed using single digital camera Canon EOS 30D and zoom lens Canon EF28-135 mm F3.5-5.6 IS USM. The test object used is a part of wing surface with a 550x550–mm^2 area, measured by capturing six images of the wing from different camera poses, with 361 selected points on each image. Note that the whole measurement including the calculation time (mainly influenced by the bundle adjustment) was actually carried out with a PIV 1.80 GHz PC within 1 minute.

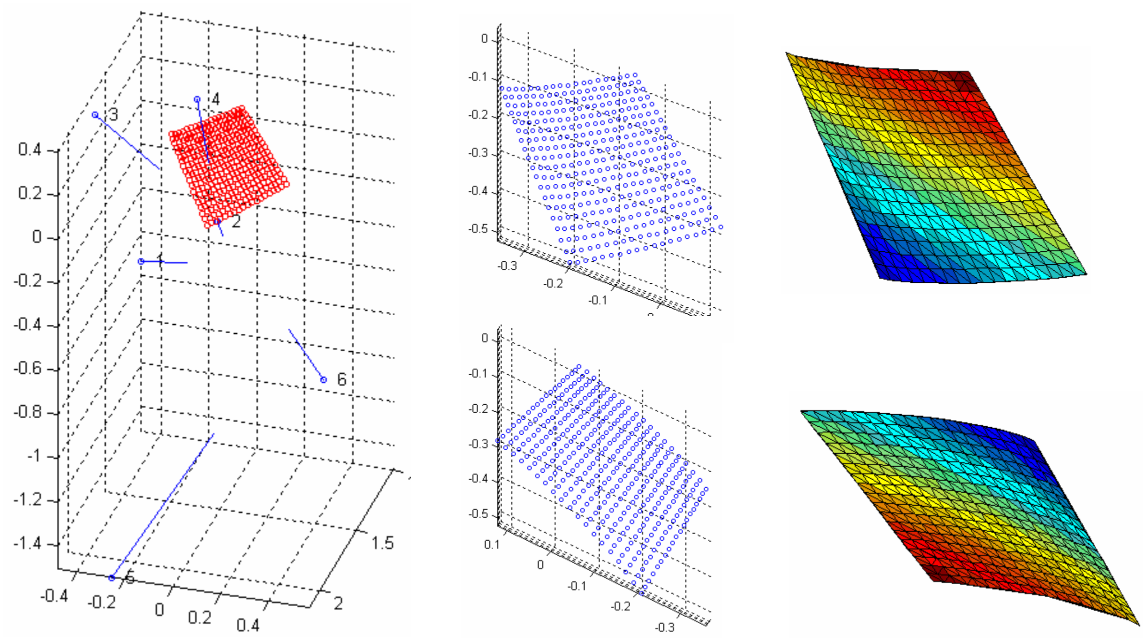
Figure 4a shows the mixing of six single views to form the total 3-D point cloud with camera poses (Figure 4b), without camera poses (Figure 4c) and the wing surface reconstructed from the obtained point cloud with Delaunay triangles (Figure 4d). Figure 5 represents the errors generated in each camera, which is the reprojection errors of the measured image points compared to the selected image points. The performance of the reconstruction is summarized in Table 1.

The experimental results show that:

1. The realization of in-process self-calibrating systems is possible without the use of calibration equipment.
2. The obtained measurement error is about 10^{-3} of the surface size with processing time of several seconds.



(a)



(b) (c) (d)

Figure 4. Reconstruction of a wing surface. (a) Six images of the wing captured at different positions and orientations. (b) Reconstructed point cloud and camera poses of the wing. (c) Point cloud of the wing. (d) Reconstructed surface from point cloud by Delauney triangles.

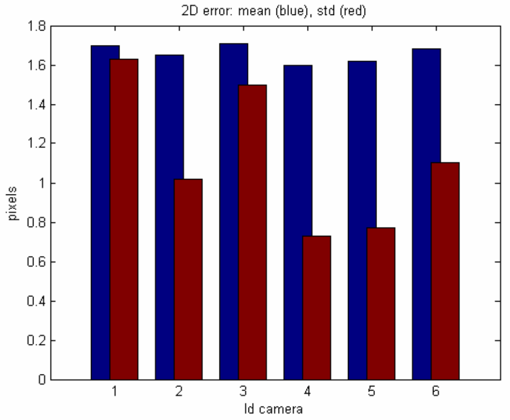


Figure 5. Average reprojection error (blue) and standard deviation (red) in each image of the wing sequence.

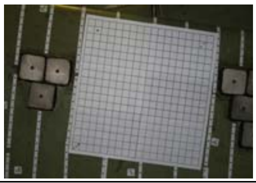
Scene name	WING					
Point detection	Manual					
Number of points	361					
Processing time [s]	15.40					
Error (before BA) [px]	2.42					
Error (after BA) [px]	1.66 (~ 0.83 mm)					
Image No. [1728 x 1152]	1	2	3	4	5	6
Mean error [px]	1.70	1.65	1.71	1.60	1.62	1.68
Std. error [px]	1.63	1.02	1.50	0.73	0.77	1.10

Table 1. Reconstructed wing surface data.

4. Conclusions

Self-calibrating surface reconstruction algorithm was proposed based on the concepts of photogrammetry. It was shown necessary to have a self-calibrating measurement for the purpose of reconstruction of various surface sizes. To achieved this, a matrix factorization operation was introduced.

The advantages of such measurement technique are versatility, simplicity and inexpensiveness. First measurements verified the concept of the proposed self-calibrating technique. The uncertainty of the measurements obtained was 1 part in 1000 of the surface size.

Using the described measurement procedure and equipment, a code package written in Matlab® is provided which could be more developed to cover large number of applications.

Acknowledgments

This research is funded by JICA – AUN/SEED-Net Collaborative Research Project 2006 – 2008. JICA – AUN/SEED-Net scholarship for the first author is also gratefully acknowledged.

References

[1] Hartley, R.I. and Zisserman, A. (2000), *Multiple View Geometry in Computer Vision*, Cambridge University Press.

[2] Tomasi, C. and Kanade, T. (1992), Shape and motion from image streams under orthography: A factorization method, *International Journal of Computer Vision*, Vol. **2**, 137-154.

[3] Sturm, P. and Triggs, P. (1997), A factorization based algorithm for multi-image projective structure and motion, In *Proc. 4th European Conference on Computer Vision*.

[4] Jacobs, D. (1997), Linear fitting with missing data: Applications to structure from motion and to characterizing intensity images, *IEEE Conference on Computer Vision and Pattern Recognition*, 206-212.

[5] Fischler, M.A. and Bolles, R.C. (1981), RANdom SAmple Consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. Assoc. Comp. Mach.*, Vol. **24**, 382-395.

[6] Svoboda, T., Martinec, D. and Pajdla, T. (2005), A convenient multi-camera self-calibration for virtual environments, *PRESENCE: Teleoperators and Virtual Environments*, Vol. **14**, 407-422.

[7] Triggs, B., McLauchlan, P., Hartley, R.I. and Fitzgibbon, A. (1999), Bundle adjustment – A modern synthesis, *International Workshop on Vision Algorith: Theory and Practice*, 298-372.

[8] Lourakis, I.A. and Argyros, A. (2004), The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm, Technical report, FORTH-ICS/TR-340.