Reconstruction of 3-D space structure with rotational imaging system

Michael Hild and Kazuyuki Nishijima
Osaka Electro-Communication University, Graduate School of Engineering
Neyagawa, Osaka 572-8530, Japan
hild@hilab.osakac.ac.jp

Abstract

We examine whether the accuracy of 3-D space point reconstruction from image pairs can be improved by using rotational imaging. The necessary mechanism and mathematical foundations of such a system are described together with the control algorithm needed for achieving optimal reconstruction results.

1 Introduction

In recent years it has become apparent that for obtaining reliable solutions of many recognition and scene understanding problems it is necessary to acquire 3-dimensional data of objects and scenes. One of the most popular approaches to acquiring such data has been 3-D space point reconstruction from stereo vision. [1]. However, 3-D space point reconstruction from conventional stereo vision employing fixed spatial configurations of cameras has been plagued by several problems: (a) there is an exponential increase of depth estimation error as the distance between space points and the camera increases (this problem is attributed to fixed baseline distance) [2]; (b) many mismatches occur during point correspondence computations especially when there are multiple similar feature points on the epipolar lines [3, 4]; (c) periodic structures in the images often cause mis-matches; (d) occlusions are hard to detect and can cause mis-matches [5], and (e) with 2-camera stereo, only those edge points can be matched reliably that have a gradient vector roughly pointing in the direction of the baseline.

In this paper we describe a new attempt to solving the problem of metric 3-D space point reconstruction by passive imaging. We use a single camera that can be rotated around two space axes not passing through the camera's optical center and reconstruct 3-D space points from stereo image pairs acquired by the camera while pointing in different spatial directions. The goal of this study is to provide an answer to the question of whether one can solve the above mentioned problems by using this mechanism, i.e. whether it is possible to achieve lower depth estimation error, increased reliability of computed matches in the presence of occlusions, periodic scene structures and multiple similar image points on the epipolar line. We assume the scene to be static, and we require that the system can carry out *omnidirectional reconstruction* of the surrounding environment's 3-D space points.

2 Motivation for rotational imaging based 3-D reconstruction

The problem of *exponential increase of depth estimation error* can be alleviated to some extend by introducing multiple baselines with different baseline distances into stereo system design [6]. However, construction of such systems is expensive and there is a tendency that they become bulky and difficult to handle. In contrast, using just one camera that is rotated around a central axis is another way of constructing stereo systems with variable baseline distances. Such a system even allows to adjust the cameras to achieve *optimal baseline distance*.

The problem of *mis-matches due to multiple similar feature points* can be solved by using a rotational camera system in which as the first step a reliable preliminary match of a feature point is obtained by rotating the camera only slightly (while tracking the feature point as the camera rotates). Then in the second step one predicts the point's location on the sensor plane after a larger rotation will have taken place. In a third step the feature point can then finally be matched in the vicinity of the predicted location, effectively avoiding mis-matches.

In the same way, *periodic image structures* can be uniquely matched by a rotational imaging system.

When using a rotational imaging system, space points that are *visible in one view and occluded in another* can effectively be dealt with: Firstly, occlusions can be easily detected because if a match cannot be confirmed in the vicinity of the location it was predicted to exist, the space point can be judged as being occluded in the second frame. And

secondly, it is possible to match points that initially are occluded by rotating the camera only as much as is required to let the point stay visible and only then carry out the match. In this case we may not be able to reconstruct the space point's 3-D location with an optimal baseline distance, but at least we can obtain the match.

The problem of obtaining matches for all extractable feature points, not only for those that have a gradient vector pointing in the direction of the baseline, can be solved in a rotational imaging system by providing for *two degrees of rotational freedom*, with one rotation occurring around the horizontal axis and another one around the vertical axis. In this case, the system's control unit must decide which of the two rotations is most appropriate for a given feature point.

In summary, 3-D space point reconstruction using rotational imaging holds the promise of providing solutions for most of the major problems exhibited by conventional stereo based reconstruction systems.

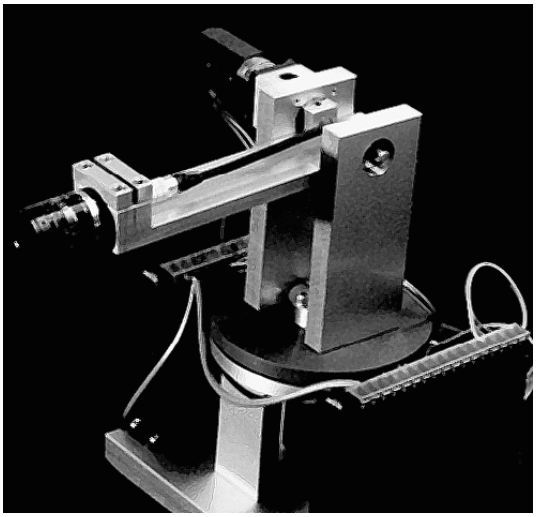


Figure 1. Prototype mechanism for rotational imaging

3 Point reconstruction by rotational imaging

3.1 Principal considerations and mechanism

In conventional stereo systems 3-D space points are reconstructed by first computing the *visual disparity* between two images of the same 3-D space point taken at different space locations and then use it for reconstruction. In the context of rotational imaging for 3-D space point reconstruction, the system generates visual disparities between two point images by rotating the camera. Since the center of the camera's optical system must be moved, the center of rotation and the center of the camera's optical system must not coincide.

A mechanism for realizing a rotational imaging system satisfying this condition and the requirement of having two degrees of rotational freedom is shown in Fig.1. The largest possible rotation angle around the (vertical) Y_c-axis (i.e. the pan angle interval) is $360°$, and the largest possible rotation angle around the (horizontal) X_c-axis (i.e. the tilt angle interval) is $270°$, making the system fully omnidirectional. The two axes of rotation intersect at the *center of rotation*, which is used as the coordinate origin of the system. The optical center of the camera is placed at a finite distance from the origin. The angular resolution of both rotations is high (approximately $0.01°$), and the rotations can be carried out precisely by means of high-precision stepping motors. Rotational slip is assumed to be close to zero due to specially designed gears.

The *world coordinate system* $(\mathbf{C}, X, Y, Z)$ is assumed to have its origin at the center of rotation $\mathbf{C}$. The Y-axis is vertical, and the X and Z axes span a horizontal plane. When the camera is in *standard pose*, the X_c rotation axis coincides with the X-axis, the Y_c rotation axis coincides with Y-axis, and the camera's optical axis coincides with the Z-axis. All space points are represented in the world coordinate system.

Given this mechanism and coordinate system and assuming that the camera is in standard pose and that its optical center is located at the origin of the world coordinate system, the camera projects a space point $\tilde{\mathbf{M}} = (X, Y, Z, 1)^T$ as

$$\tilde{\mathbf{m}} = \tilde{\mathbf{P}}\tilde{\mathbf{M}}, \tag{1}$$

where $\tilde{\mathbf{m}} = (U, V, S)^T$ is a column vector in homogeneous coordinates and S is a scale factor equal to Z, and $\tilde{\mathbf{P}}$ is the 3x4 projection matrix in homogeneous coordinates. The Euclidean coordinates of the space point's projection onto the image (sensor) plane of the camera are given as $u = U/Z, v = V/Z$.

Since the camera's optical center $\mathbf{c}$ and the center of rotation $\mathbf{C}$ must not coincide, the camera has to be translated along the Z-axis by some distance Z_c. This is done by defining a translation matrix $\tilde{\mathbf{T}}$ in projective coordinates and concatenating it to the projection matrix.

When the 3-D position of a space point is to be determined, the camera first rotates around the X_c rotation axis by angle β in oder to adjust the orientation of the Y_c-axis, then it rotates clockwise by some angle α_{cw} around the Y_c-axis to position CP_1, takes an image of the space point $\mathbf{M}$, then it rotates counterclockwise by some angle α_{ccw} around the Y_c-axis to position CP_2, and takes the second image of the space point. (Similarly, the same can be done with rotations in up- and downwards directions, where X_c-axis and Y_c-axis are exchanged). The two rotation angles usually satisfy the condition $\alpha_{cw} < \alpha_{ccw}$, and the rotations must be

such that the point to be estimated is visible from both camera positions. These rotations are expressed by two matrices $\tilde{\mathbf{R}}_y^{cw}$ and $\tilde{\mathbf{R}}_y^{ccw}$ given in homogeneous coordinates. The projection of the space point when the camera is at position CP_1 is thus given by

$$\tilde{\mathbf{m}}_1 = \tilde{\mathbf{P}}\tilde{\mathbf{T}}\tilde{\mathbf{R}}_x\tilde{\mathbf{R}}_y^{cw}\tilde{\mathbf{M}} \tag{2}$$

and the one at position CP_2 is given by

$$\tilde{\mathbf{m}}_2 = \tilde{\mathbf{P}}\tilde{\mathbf{T}}\tilde{\mathbf{R}}_x\tilde{\mathbf{R}}_y^{cw}\tilde{\mathbf{R}}_y^{ccw}\tilde{\mathbf{M}} \tag{3}$$

If translations and rotations are combined with the projection matrices for the two camera positions, the resulting matrices are given by

$$\tilde{\mathbf{P}}_1 = \tilde{\mathbf{P}}\tilde{\mathbf{T}}\tilde{\mathbf{R}}_x\tilde{\mathbf{R}}_y^{cw} \tag{4}$$

and

$$\tilde{\mathbf{P}}_2 = \tilde{\mathbf{P}}\tilde{\mathbf{T}}\tilde{\mathbf{R}}_x\tilde{\mathbf{R}}_y^{cw}\tilde{\mathbf{R}}_y^{ccw}, \tag{5}$$

and the two projections can concisely be expressed as

$$\tilde{\mathbf{m}}_1 = \tilde{\mathbf{P}}_1\tilde{\mathbf{M}}, \qquad \tilde{\mathbf{m}}_2 = \tilde{\mathbf{P}}_2\tilde{\mathbf{M}}. \tag{6}$$

When matching the image of a space point to its corresponding image, the epipolar line is useful as a constraint. It is computed a follows: First, the 3-D world coordinates $\tilde{\mathbf{M}}_1$ of point projection $\tilde{\mathbf{m}}_1$ (on the sensor plane) are computed, then the optical center of the camera at position CP_1 as given by $\tilde{\mathbf{C}}_1 = \tilde{\mathbf{T}}\tilde{\mathbf{R}}_y^{ccw}(0,0,0,1)^T$ is computed, and then both of these points are projected onto the sensor plane of the camera at position CP_2, resulting in points $\tilde{\mathbf{m}}_{1,2}$ and $\tilde{\mathbf{e}}_2$. The line through $\tilde{\mathbf{m}}_{1,2}$ and $\tilde{\mathbf{e}}_2$ is the epipolar line used for matching. In rotational imaging, epipolar lines are not parallel to the x-axis of the image coordinate system, but if the rotation angle between two views is small, they can be considered as being approximately parallel to the x-axis.

3.2 Optical system modeling

The accuracy of reconstruction of 3-D space points strongly depends on accurate modeling of all system components, including the optical system. Since reconstruction by rotational imaging involves camera rotations around an axis that does not pass through the camera's optical center, the spatial locations of the optical centers have to be determined accurately after each rotation. For achieving this, the pinhole model commonly used for representing the optical system for stereo imaging is no longer adequate. Instead, we represent the optical system as a *thick lens* for which in place of the pinhole two nodal points N' and N'' are used [7].

The imaging light rays are assumed to pass through these two nodal points as is shown in Fig.2 in the context of the rotational imaging system, with the camera being depicted in standard pose. The figure shows how a space point $\mathbf{M}_1 = (X_1, 0, Z_1)^T$ in world coordinates is projected onto the sensor plane at $\mathbf{m}_1 = (u_1, 0)^T$ (in image coordinates). The distance between the two nodal points is denoted H, the distance between space point $\mathbf{M}_1$ and anterior nodal point N' is denoted g, and the distance between posterior nodal point N'' and the sensor plane is denoted by b.

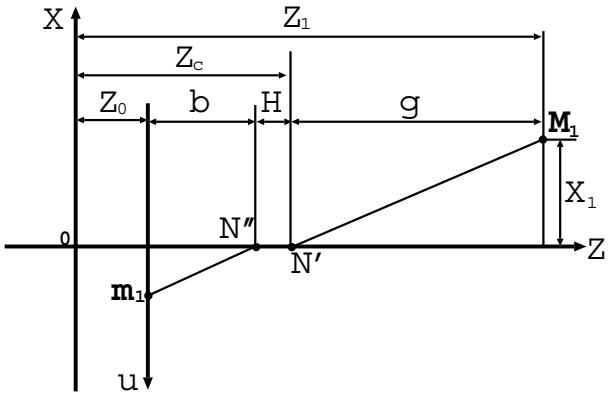


Figure 2. Optics

The two nodal points split the projecting ray into a anterior ray and a posterior ray, both of which have the same direction. Between the nodal points, the ray is assumed to run along the optical axis of the system (i.e. we actually assume a "pin-pipe" in place of the commonly assumed pin-hole). Although the real optical center of the lens is located somewhere between the two nodal points, in this idealized model of imaging we can assume that the optical center of the system is located at anterior nodal point N'. Thus, the distance between the center of rotation and the optical center is given by

$$Z_c = H + b + Z_0 \tag{7}$$

With this geometric set-up the projection matrix of the camera becomes

$$\tilde{\mathbf{P}} = \begin{pmatrix} b & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{8}$$

We assume that the lens' focal length f and distances g and b are related by Gauss's lens equation

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}. \tag{9}$$

Provided that focal length f and distance b are fixed (as is usually the case), only space points with distance g to the plane containing anterior nodal point N' that satisfy Gauss' lens equation are projected onto the sensor plane with a unblurred image. Consequently, locations of point images not

satisfying Gauss' lens equation will be estimated with some small error. We neglect this error.

3.3 Metric reconstruction of space points

Metric reconstruction of a space point $\mathbf{M}$ of which we know only its two projections $\mathbf{m}_1$ and $\mathbf{m}_2$ at camera positions CP_1 and CP_2 can be carried out by several methods. Probably the most often used method consists of formulating this problem as a least squares estimation problem [8]. We used a different method which is based on the intersection of two projection rays.

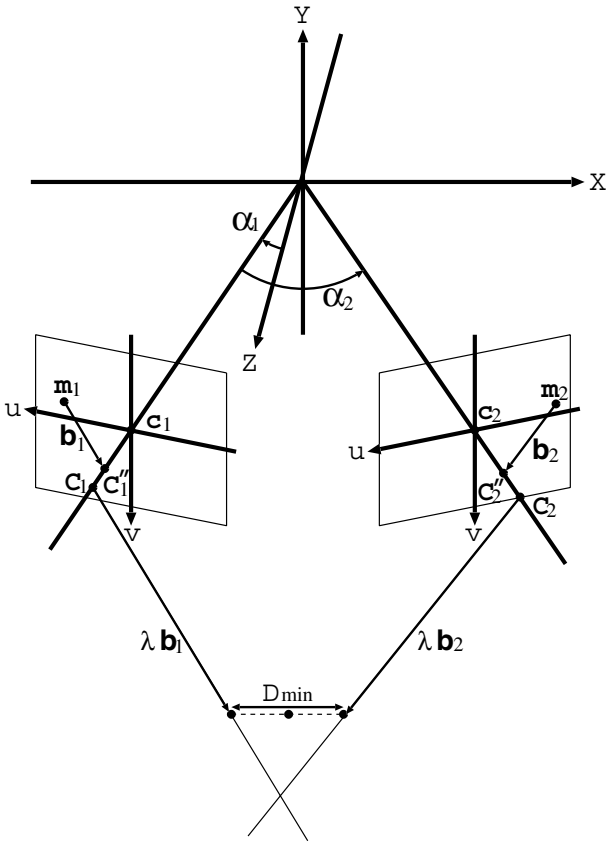


Figure 3. Reconstruction

In this method, 3-D space points are reconstructed from the point's images $\mathbf{m}_1$ and $\mathbf{m}_2$ by projecting the two images backward through the two nodal points N' and N'' and then determining the intersection of the two projecting rays. This situation is shown in Fig.3. The locations of the anterior nodal points are indicated by position vectors $\mathbf{C}_1$ and $\mathbf{C}_2$ (which are centers of the optical system), and those of the posterior nodal points are indicated by $\mathbf{C}_1''$ and $\mathbf{C}_2''$. The two projecting rays $\mathbf{b}_1$ and $\mathbf{b}_2$ can be described as equations of straight lines given in parametric vector form:

$$\mathbf{g}_1(\lambda_1) = \mathbf{C}_1 + \lambda_1 \cdot \mathbf{b}_1, \qquad \mathbf{g}_2(\lambda_2) = \mathbf{C}_2 + \lambda_2 \cdot \mathbf{b}_2 \quad (10)$$

where $\mathbf{b}_1$, $\mathbf{b}_2$ are vectors connecting the space point images $\mathbf{m}_1$ and $\mathbf{m}_2$ with the two posterior nodal points. Ideally, these two lines would intersect each other at the sought space point $\mathbf{M}$, but due to the discrete nature of the image sensor's sampling array, estimation errors and other inaccuracies in system parameters such exact intersections rarely exist. Therefore in practice the best estimate of $\mathbf{M}$ will be given by two points on the two lines whose distance is minimal. Assuming that this happens when $\lambda_1 = \lambda_1^*$ and $\lambda_2 = \lambda_2^*$, space point $\mathbf{M}$ is determined as

$$\mathbf{M} = (\mathbf{g}_1(\lambda_1^*) + \mathbf{g}_2(\lambda_2^*))/2 \quad (11)$$

3.4 Stereo image matching based on prediction of match locations

In the proposed rotational imaging system a 3-D space point is reconstructed from two images of the point; these images are acquired while the camera is pointing in two different spatial directions. Assuming a *base direction* in which the part of the scene to be reconstructed is located, the camera has to carry out at least two rotations for acquiring these two images. A control algorithm for acquiring images that are taken in the two "optimal" directions is described in this section.

In order to obtain the most accurate reconstruction, the rotation angle between the two camera orientions should be as large as possible. (Evidence for this assumption was found through simulations.) Consequently, the camera should be rotated in a way that a space point to be reconstructed would be projected onto the sensor plane near opposite rims of the sensor's area. However, the disparities will be large in this case, which is equivalent to a high risk of mis-matches. In order to nontheless be able to achieve good reconstruction accuracy as well as a low rate of mismatches, we adopt an imaging control strategy that obtains the reconstruction in three steps:

1. Only a small-angle rotation guaranteeing high matching reliability is used for preliminary, low-accuracy reconstruction of the space point, then
2. optimal rotation angles are predicted together with search intervals near the rims of the two images in which the matches are likely to occur, and finally
3. accurate reconstruction is carried out based on the predicted match locations using the optimized, wide baseline distance of the two camera positions.

The search intervals used in the control algorithm are derived from knowledge of the near and far limits of the optical system's depth-of-focus range, as well as near and far

limits of the depth error range estimated for each preliminarily reconstructed space point.

Although matching itself is carried out by the sum-of-absolute-differences (SAD) method within a small window region placed on the gray-scale images, these matches are computed only at locations where edge points occur. Edge points must satisfy two conditions: (1) the magnitude of the gray-scale gradient vector must exceed a pre-determined threshold, and (2) the direction of the gradient vector most be roughly aligned with the direction of the epipolar line on each image. The locations of edge points are determined to sub-pixel accuracy based on computing zero-crossings of the interpolated Laplacian-of-Gaussian of the gray-scale image.

Although for simplicity reasons the above control algorithm is stated for the case of reconstruction of only a single edge point, the algorithm actually used is more sophisticated since it carries out the matching and reconstruction for *groups of similar feature points*. This allows to significantly reduce the total number of camera rotations needed for matching, making the system more efficient.

4 Experimental results

The mathematical description presented in the preceding section allows to examine the fundamental properties of 3-D space point reconstruction based on rotational imaging by simulating the effects that errors in various system parameters may have. From such simulation results we have concluded that the rotational imaging system should be designed to have a long distance Z_0, a long focal length f, and the rotation angles always should be adjusted to the largest possible values, but the most crucial choice would be the device for measuring rotation angles – its accuracy must be very high.

We have built a pan-tilt mechanism for rotating a monochrome CCD camera with a 646x485 pixels sensor and a lens with focal length of approx. 50 mm (refer to Fig.1(a)) and used it for experiments. The stepping motors built into the mechanism have an angular resolution of 0.01° and mechanical backlash is prevented by a harmonic drive gear. For obtaining the following results we used mainly target objects that were placed approximately 1500 mm from the mechanism's center of rotation. The results are as follows:

Accuracy of general stereo image matching: A bunch of artificial flowers made of textile was placed at a distance of approximately 1500 mm from the origin of the world coordinate system under usual laboratory lighting and had the rotational imaging system automatically obtain feature points and their matches. Among them, mis-matches were detected manually by sorting out those feature points whose 3-D reconstruction was obviously wrong. The number of such mis-matches out of 1547 extracted feature points was 79, which is equivalent to an error rate of 5.11%. In a separate experiment a small toy bird made of plastic and painted on the surface was imaged. The number of matched points was 434, among which 11 points (2.53%) were obvious errors. It should be noted that these matches were non-periodic pattern matches.

Accuracy of matching in the presence of occlusions For this experiment, one planar board with well-identifiable feature points was placed at a distance of 1490 mm from the origin of the world coordinate system ($Z = 1490mm$) and another planar board was placed at a distance of $Z = 990mm$ in such a way that some of the feature points on the board in the back were not visible from one of the two camera positions. When under these conditions the system automatically estimated the 3-D locations of extracted feature points, it did not return any mis-matches.

Accuracy of matching in the presence of periodic patterns For this experiment, planar chess board patterns were placed at location $Z = 1490mm$, with the boards' orientation being orthogonal to the Z-axis. The edge points of the chess patterns served as feature points. The width of the chess pattern squares was varied during the experiment. The results are summarized in Table 1 where n is the number of extracted feature points, m is the number of mismatched points, and the right-most column shows the error percentage. The system is able to handle periodic patterns flawlessly up to a certain period below which the prediction based-control algorithm begins making mistakes.

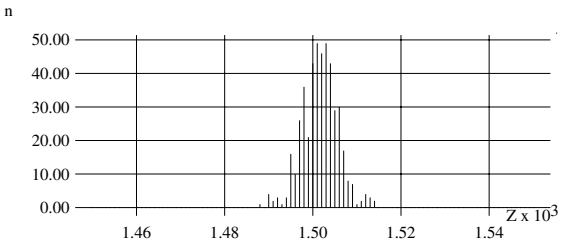
Table 1. Mis-matches due to periodicity of scene.

square width [mm]	*n*	*m*	%
44.5	1381	0	0.00
31.6	1362	0	0.00
19.2	2179	0	0.00
6.3	3955	0	0.00
4.0	2800	245	8.75
3.5	3931	293	7.45

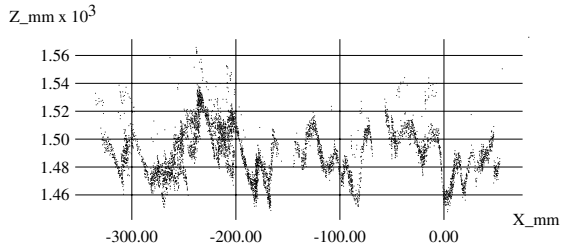
Precision of reconstructed 3-D coordinates: Fig.4(a) shows the histogram of distance estimation results obtained for a planar target object containing albedo edge points that had all the same Z-world coordinate values ($Z = 1490mm$) and where aligned vertically. The average of the estimates was $1501.68mm$ (i.e. an absolute average error of $11.68mm$ or 0.78%), and the standard deviation was $16.12mm$.

Errors occurring when the 3-D locations of horizontally aligned feature points on the same test board are estimated

are shown in Fig.4(b). The error appears to have a quasi-periodic footprint. This indicates that the error is mainly attributable to the error of the stepping motor's rotation angle which is specified by the maker as about $2'$, having an almost periodic characteristic. It can be shown through simulations that the stepping motor's rotation angle error of about $2'$ causes a distance error of about the magnitude observed.



(a) Histogram of depth estimates for planar target



(b) Depth estimates vs. x-location

Figure 4. Errors of depth measurement

From the above results it can be concluded that rotational imaging systems are able to avoid match errors in the presence of occlusion and periodic patterns. The precision of 3-D reconstruction (mainly the distance estimate) is with about 1% quite acceptable for many applications, but if it would be possible to improve the measurement of the rotation angle, much better results can be expected.

5 Conclusion

In this paper we have examined whether the accuracy of 3-D space point reconstruction from image pairs can be improved by using rotational imaging. We described the mathematical foundations of such a system together with the control algorithm needed for achieving optimal reconstruction results. We examined the properties of this method based on quantitative simulation results and the results of experiments carried out with a laboratory implementation of the rotational imaging mechanism and processing software. These results show that the rotational imaging approach leads to low matching errors and does not cause mismatches in the presence of occlusions and periodic image patterns. 3-D reconstruction accuracy was very good for X, Y coordinate values, and the accuracy of the range estimate (Z coordinate) was approximately the same as in existing systems. We found that the result for range estimation is due to rotation angle inaccuracies of the stepping motors used in the mechanism, and we believe that much better results can be achieved if improved rotation angle measurement methods are used.

References

[1] N. Yokoya, T. Shakunaga, M. Kanbara, Passive Range Sensing Techniques: Depth from Images, IEICE Trans. Inf.& Syst., Vol.E82-D, No.3, 523-533 (1999)

[2] A. Verri, V. Torre, Absolute depth estimates in stereopsis, Opt. Soc. Amer., Vol.3, No.3, 297-299 (1986)

[3] G. Wei, W. Bauer, G. Hirzinger, Intensity- and Gradient-Based Stereo Matching Using Hierarchical Gaussian Basis Functions, IEEE Trans. PAMI, Vol.20, No.11, 1143-1159 (1998)

[4] S. B. Marapane, M. M. Trivedi, Multi-Primitive (MPH) Stereo Analysis, IEEE Trans. PAMI, Vol.16, No.3, 227-240 (1994)

[5] D. Scharstein, View Synthesis Using Stereo Vision, Lectures in Computer Science 1583, Springer-Verlag, Berlin (1999)

[6] M. Okutomi, T. Kanade, A Multiple-Baseline Stereo, IEEE Trans. PAMI, Vol.15, No.4, 353-363 (1993)

[7] F. A. Jenkins, H. E. White, Fundamentals of Optics (Fourth Edition)), McGraw-Hill International Book Comp, Singapore (1981)

[8] O. Faugeras, Three-Dimensional Computer Vision, The MIT Press, Cambridge (1993)

Acknowledgement

This work has been supported by the Ministry of Education, Science, Sports, and Culture in Japan under a Grant-in-Aid for Scientific Research No. 11680404.