Journal of Visualization, Vol. 6, No. 2 (2003) 155-164

3D Shape Reconstruction from Synthetic Images Captured by a Rotating Periscope System with a Single Focal Direction

Someya, S.*[1], Okamoto, K.*[2] and Tanaka, G.*[2]

*1 Research Institute of Energy Utilization, AIST,
Tsukuba, Ibaraki, 305-8564, Japan.
Tel:+81-298-61-7247 / Fax:+81-298-51-7523
E-mail: s.someya@aist.go.jp
*2 University of Tokyo, Ibaraki, Japan.

Received 12 June 2002
Revised 9 September 2002

***Abstract*:** A three-dimensional shape measurement system with only one camera and a rotating periscope is proposed. Many images are captured at every few degrees of periscope rotation. The distance along the optical axis of the camera from the focal point of the camera system to a certain point along the object surface is inversely proportional to the parallax. In this system, a circular streak is obtained by tracing a point on each image using the cross-correlation along the direction of periscope shift. The parallax can be calculated from the radius of the streak. Applying this method to shape measurement of the human face from synthetic images, the images were well reconstructed and the proposed system may be adaptable to a portable 3D shape acquisition system.

***Keywords*:** Image processing, Sensor, Three-dimensional measurement, Optical flow.

1. Introduction

Three-dimensional shape reconstruction techniques have been developing rapidly in recent years. These techniques are used in applications such as CAD, industry prototyping, multimedia databases, and virtual reality applications. In addition, the remarkable popularization of personal computers has familiarized digital images, such as 3D computer graphics, which are often used in movies and computer games. Although making 3D CG pictures is complex, a much simpler 3D measurement system which is also useful in 3D information network systems via internet should be developed.

3D shape reconstruction techniques are classified into two approaches, i.e., passive systems and active systems. Active systems include contact and non-contact methods. Non-contact active systems adopt a single or a stereo camera, with rotating objects, that illuminates a characteristic optical pattern, using magneto optical sensors. These systems provide 3D shape measurement with good precision. J.Y. Zheng and A. Murata, (2000), proposed an active system, involving circling an object with ring-shaped light sources which was strongly fixed. This system can deal with even

specular surfaces. However, the applicability of any active system is limited due to its own complexity and the size of the system.

Passive systems, stereo vision systems, are relatively free from limits in usage and applicability, but require rather precise calibration of the parameters of the cameras. The multiple-vision system requires a longer disparity range, although there is a tradeoff between precision (disparity) and correctness in pattern matching. The passive system also requires high texture at the surface, similar brightness (no shadow), as well as other requirements. Another large problem with the stereo vision system is occlusion. Occlusion is a critical and difficult phenomenon to be dealt with by stereo algorithms. M. Okutomi and T. Kanade, (1993) adopted a multiple baseline stereo system to clear the tradeoff problem using many pairs of stereo images with small disparity and/or with large disparity. C.L. Zitnick and T. Kanade, (2000), proposed a stereo algorithm for obtaining disparity maps with occlusions explicitly detected, with assumptions of uniqueness and continuity. Thus, a lot of research has been carried out. However, most of these previous systems are large and complex. At least two cameras with sufficient baseline are needed, i.e., distance between each other, but this may be not suitable for family use.

The authors could not find a small system with a simple calibration, which is applicable to objects with less dependency on the surface conditions of texture, reflection, shadow and so on. Most previous methods have pursued a precise 3D measurement system, for which there is no simple and easy system. Therefore, we present a small system for 3D shape measurement in this paper. The system is composed of a single camera and rotating periscope, which is an active, non-contact system which has the merits of the multiple-baseline method and the optical flow system.

2. System and Method

2.1 Basic Concept of the System

In stereo matching, the disparity, r, which is the displacement between the corresponding points of two images, is usually measured. As shown in Fig. 1, the disparity is related to the distance, z, by the equation $z = f R / r$, where R and f are the baseline and the focal length, respectively.

Figure 2 shows a schematic diagram of our system. A mirror is placed in front of the camera at an angle of $\pi/4$. Another mirror is placed parallel to the first mirror separated by an interval R. This periscope is rotated coaxially with the fixed camera. Several pairs of stereo images are captured at small angular intervals, e.g., $2\pi/180$ in the present paper. Equivalent images can also be captured by rotating camera, while the periscope-type mirror enables the system to be downsized. The distance,

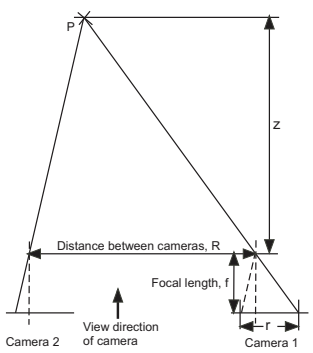


Fig. 1. Basic concept of stereo vision.

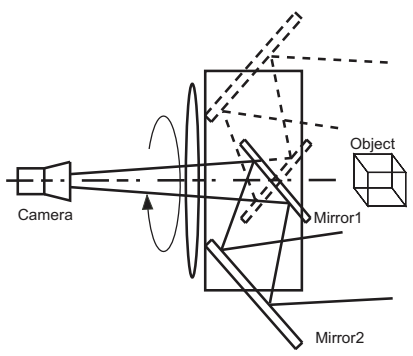


Fig. 2. Schematic diagram of the proposed system.

parallel to the axis of rotation of the periscope, from the focal point of the camera system to the corresponding point at the object surface is inversely proportional to the disparity. In our system, the baseline, R, is the radius of periscope rotation, i.e., the interval between mirrors. The disparity, r, can be calculated from the radius of circular streak, which is obtained by tracing the corresponding points on each image. In our system, points at greater distances draw circular loci of smaller radii.

One advantage of the proposed system is the reduction in size associated with the use of one camera. In the present study, the angle of the mirror in the proposed system is fixed at $\pi/4$. However, the angle can be made variable through future studies, and the system can be applied to various object despite the compactness of the system.

2.2 Overview of Algorithm

Adopting small movement of the camera makes the matching problem quite easy, which results in a short distance between corresponding points for pairs of consecutive images. So the results may include the integrated noises or errors from the periscope motion and/or from individual image analyses. The advantages of multiple pairs of consecutive images are much larger. Every captured image is related to other images by applying a geometrical constraint according to the unique motion of the periscope. Various types of averaging techniques may be adopted in the case of multi-baseline methods, not only the disparities are equal among the consecutive images but also the direction of motion flow vector in each image pair shifts over a certain angle in our system. So, by fitting the locus with a circle, constraining the displacement and the direction, it is relatively easy to remove errors from the tracking.

In terms of the calibration of the camera system, the proposed system needs to know the angle of the rotating mirror and the angle of the camera view. Even if the radius of mirror rotation is unknown, the relative 3D shape of the object can be analyzed. Even in the horizontal displacement system, almost same analysis method is applicable, although there are many blind spots. In addition, the displacement in each step must be exactly known in case of the horizontal displacement system, in which the direction of camera motion is fixed. There are more parameters which must be calibrated.

Our algorithm consists of five steps:

1) Search candidate points for correspondences, i.e., vectors of optical flow, in each image pair using the cross correlation method along the direction of periscope motion. This method can be applied coarse-to-fine processing and/or overlapped interrogation window.
2) Estimate the distance, z, to each point in the “first image” initially defined arbitrarily. Then, the distances are fed back/re-converted to motion flow vectors in image pairs. These vectors are referred to herein as “predicted vectors”.
3) Re-calculate the optical-flow vectors within the narrow search conditions, near the predicted vectors. The minimum size of the search window is fixed as one.
4) Re-estimate the distances and re-convert the results to the predicted vectors.
5) Decide the final results of distances directly from the predicted vectors, but only for each point in the first image mentioned in the 2nd step.

To get the optical flow vectors, we adopted a cross-correlation method with a 2-D search-window and a 1-D search-direction. There is no concept of the method for searching correspondences by sub-pixel scale because the 1-D search-direction is not orthogonal to the grid of image pixels. Along the known direction, i.e., θ, the correspondences of the fixed size window are searched, i.e., at $(x_n,y_n)=(x_{n-1}+(m)\cos\theta, y_{n-1}+(m)\sin\theta)$, $(x_{n-1}+(m+0.01)\cos\theta, y_{n-1}+(m+0.01)\sin\theta)$, $(x_{n-1}+(m+k)\cos\theta,\ y_{n-1}+(m+k)\sin\theta)$. Here, the coordinates are those at the center of a search-window. A coarse-to-fine processing

with a fixed size window is adopted in the calculation. The averaged intensities in these search-windows are calculated by interpolating intensities with a linear or a spline-function. By setting m and k in a certain range, the analyzed distances are limited.

The second step, prediction and feedback of vectors, is most important for this algorithm, which is very effective in order to narrow the search region and decrease error-vectors. This method is more useful for the 3D shape recognition than methods with averaging and/or interpolating vectors. Steps 4 and 5 may round off results, but they are quite effective for the 3D reconstruction. In terms of the recognition of the 3D shape at least, using the predicted vectors is more adequate than the other methods of error detection of each vector based on the averaging or interpolating techniques. The efficiency of the algorithm is illustrated by applying it to synthetic images in the next section.

In order to estimate the distances of points, we compared three algorithms. In the following equations, x_c, y_c are the coordinates of the center of circular streaks, x_i, y_i are the tracked coordinates of points in images, r is the radius of circular streaks, and x_s, y_s are x_i, y_i at i=0. θ_i is the rotation angle of the periscope. θ_s is θ_i at i=0, and "$(\theta_{i+1} - \theta_i)$" is constant in this system.

i) The circular streaks are tracked counterclockwise and clockwise in images, halfway around in each direction, while they are aborted to track in the case of reduced reliability. The unknown radius, r, is calculated by the least squares method with a known rotation angle and the calculated disparity in each pair, according to equations (1) and (2). We call this algorithm the "one-freedom method" in this paper.

$$x_c = x_s + r \cdot \cos\theta_s \qquad y_c = y_s - r \cdot \sin\theta_s \tag{1}$$

$$r = \sum_i \{(x_i - x_s)(\cos\theta_s - \cos\theta_i) + (y_i - y_s)(\sin\theta_i - \sin\theta_s)\} \Big/ \sum_i 2(1 - \cos(\theta_i - \theta_s)) \tag{2}$$

ii) The points are also tracked halfway around in both directions. The radius, r, and the coordinates of the center of circular streaks are calculated by the least squares method. This algorithm, represented by equations (3) and (4), is called the "three-freedom method" in this paper.

$$x_c = \sum_i (x_i + r \cdot \cos\theta_i) \Big/ n \qquad y_c = \sum_i (y_i - r \cdot \sin\theta_i) \Big/ n \tag{3}$$

$$r = \sum_i \left\{ \left(\frac{\sum x_i}{n} - x_i \right) \cos\theta_i - \left(\frac{\sum y_i}{n} - y_i \right) \sin\theta_i \right\} \Big/ \sum_i \left(1 - \frac{\cos\theta_i}{n} \sum_i \cos\theta_i - \frac{\sin\theta_i}{n} \sum_i \sin\theta_i \right) \tag{4}$$

iii) The radius, r, is calculated from the circumference of a circle, the sum of disparity in each pair as shown in eq.(5), which is called the "direct method" herein.

$$r = \frac{1}{2\pi} \sum_j^n r_j \tag{5}$$

3. Results

In order to clarify the effectiveness of the proposed system, we use a synthetic polygon 3D data called "tiny.x", which is presented with Direct3DX software by Microsoft Corporation. One of the aims in this research is developing an easy handling 3D shape acquisition system suitable for private,

family use. So we selected images of a simulated human subject. From the polygon data, we can get 2-D distortionless images with precise camera parameters, i.e., the camera location, orientation vector and so on. In this paper, the image size is fixed at 512 x 512 pixels, and the image depth is fixed to 8 bits gray scale, for calculation. Configurable parameters are the angle of camera-view, the radius of rotation of mirror and the distance. These are set at 45°, 0.5 and 2.0, respectively. The object size, e.g., TINY's face (w x h, 120 x 200 pixels), is equivalent to a width of 0.387 and a height of 0.645 with respect to the distance, 2.0. A default number of used images, 180 images, are used to simulate the proposed system. Figures 3(a) and (b) show examples of images at angles of 0 and π, respectively. The left side image at an angle of 0 is defined as the first image mentioned in section 2-2 (step 2). The distances, z, at the points only in the left-side image are calculated. Only the area around the face (shown by a rectangle in the left-side figure, about 200 x 200 pixels) can be seen in every image through a rotation of mirror. Figure 3(c) shows the face in detail from the side in order easily to compare it with the results.

Theoretically, a 3D shape can be reconstructed from only a pair of distortionless stereo images with precise camera parameters. However, in the case of Figs. 3(a) and (b), there are occlusions, points out of sight in one image. Namely, the left cheek can be seen only in the left side image. The image resolution is not so high. Parts having large disparities have large deformations. Therefore, it is very difficult to reconstruct TINY's face three-dimensionally from these two images.

Figures 4 shows the reconstructed images using the one-freedom method. Figure 4 shows different angles of view and illumination in order to clarify sharpness and smoothness of the images. The original textures, colors, of images are directly mapped on the results. Figure 4(a) shows the

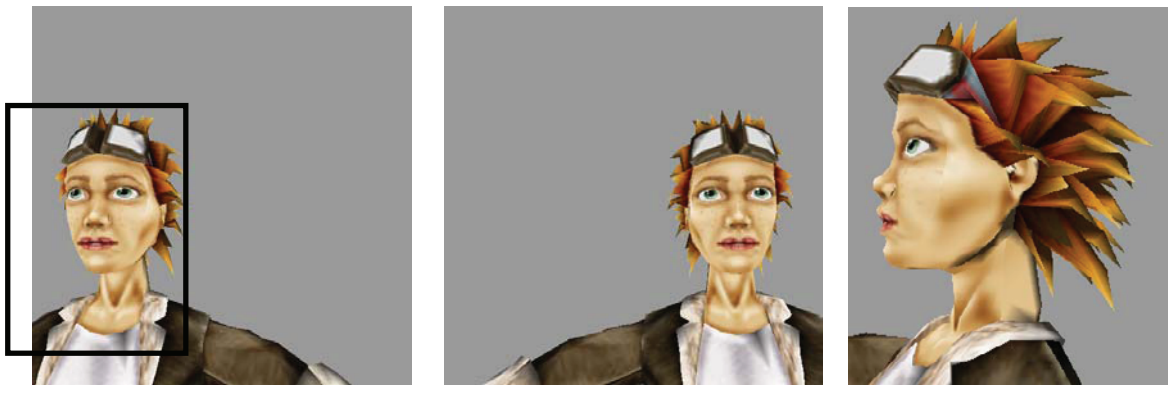


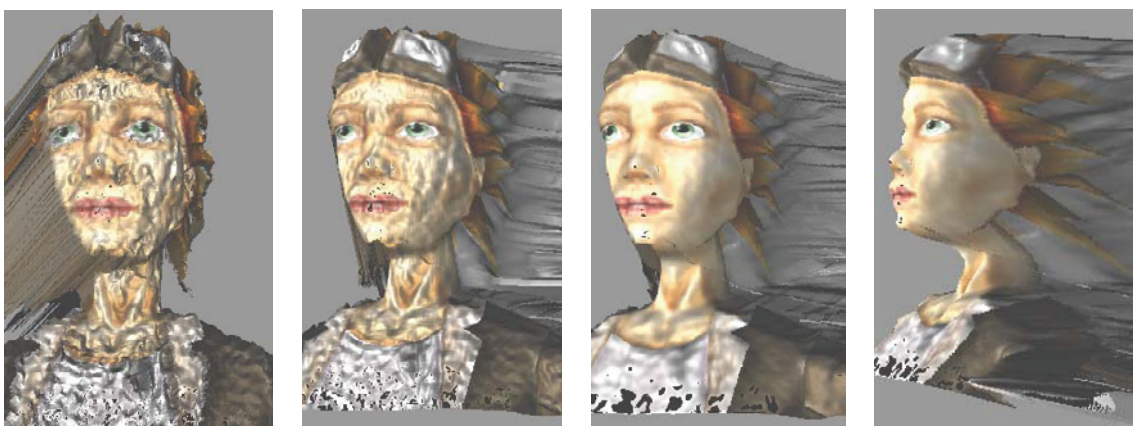
Fig. 3. Examples of examined images (Microsoft © TINY in Direct3DX) and a detailed side view.

distance of each point estimated from motion flow vectors in the 1st step, which was calculated using the 2-D cross correlation technique, for 5 x 5 pixels of the interrogation window. Correspondence of each point should be found in a known direction fixed by periscope movement; whereas finding the constraints for correspondences like epipolar lines is a large problem for many other systems of stereo vision. The pixel intensity is interpolated by a spline function. Vectors in the motion flow field are interpolated linearly. Applying a coarse-to-fine processing while changing the size of the search-window and/or overlapped interrogation window is not so effective on our system, at least in this simulation, so these algorithms are not used here. Because of the small angle interval for each image pair, there is not so much occlusion locally. In addition, the circular streaks are tracked reciprocally to be more robust with respect to global occlusions, where the points are invisible,

dependent on its view direction.

In order to estimate the effects of processes and conditions, the motion flow vectors and the shapes should be compared with the correct processes and conditions. The synthetic polygon 3D data of TINY has a correct shape, while the correspondences of grids (5x5 pixels) in all image pairs can not be decided exactly. Average values of correct motion flow vectors of points in the grid are not appropriate as the correct motion flow vectors, because a 3D shape reconstructed from these vectors becomes somewhat blunt. Therefore, we defined "correct vector" as a correct motion flow vector of a center point in a grid. The accuracies of the system are discussed using the number of successfully measured points without error (0.9995 < calculated one/correct one < 1.0005), though the number of all of tracked vectors, which includes meaningless vectors for the background image, differ case by case.
In Fig. 4(a), the shape of TINY is quite rough and rugged, especially near the hair and goggles. There are about 1000 points with no error in the reconstructed shape. From the result shown in Fig. 4(a), the predicted vector values of motion flow fields for the next step are given, which are quite useful for limiting the search area and for decreasing erroneous correspondences. Many researchers used to iterate the calculation of cross correlation using the motion flow vectors in the last step between only two images, combined with a coarse-to-fine processing. Such an algorithm may become sensitive to errors, at least in the case of low texture of images and in the case of the lack of precise calibration of the camera. On the other hand, in our case, a motion flow vector for a certain point in each pair of images should represent the same depth of the point, $z(=fR/r)$. Disparity, r, is represented by the sum of motion flow vectors, r_j, as $r=2r_j$ (e.g., in case of the direct method). The adjustments by depth are more effective than those by motion flow vectors in the last step, although this may make the final shape somewhat dull.

Figure 4(b) shows the re-calculated result, which is still rugged, but the goggles and hair are reconstructed slightly better than those in Fig. 4(a). The number of successfully measured points without error is about 1460, thus predicted vectors functions are applied quite well. The distance at each point obtained from the result in Fig. 4(b) is converted into the next predicted vector value. Then, distances are estimated again from the predicted vectors without another calculation of cross correlation, considering the predicted vectors as the calculated motion flow vectors. Figures 4(c) and (d) show the final reconstructed images from different angles of view. Thus, the 3D shape is reconstructed in detail, near the lips, the nose and the eyebrows and so on. The number of vectors tracked without error is about 1590 at this final step, which is a slight increase, even in the last process, for rounding off the reconstructed feature.



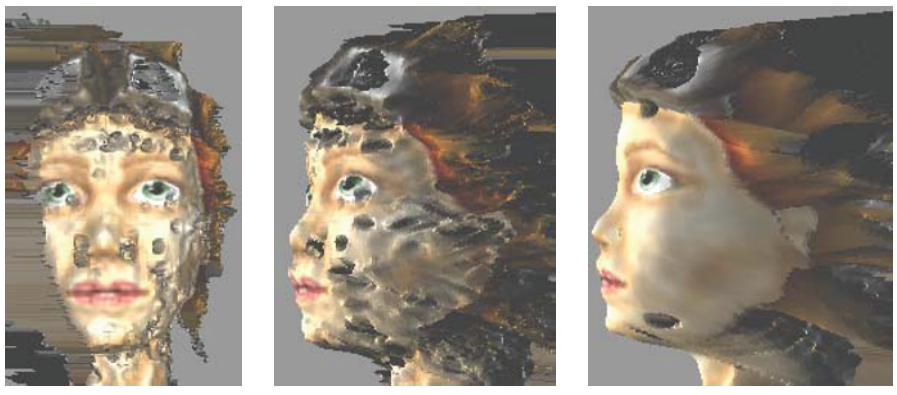
(a) (b) (c) (d)

Fig. 4. Reconstructed images for each step.

Table 1 summarizes the effects of parameters by the number of points successfully measured without error. The result at each step becomes worse in the case that the intensities in pixels are interpolated linearly and becomes slightly worse in the case that the spline interpolation is applied for the vectors in motion flow fields. Figures 5-7 show reconstructed results from 45, 90 and 250 images. As shown in Table 1, the number of correct points, which is reconstructed from 45 images, becomes the largest at the 1st step. From 90 images, the number is the largest at the 3rd step. However, the reconstructed images are very rough and rugged in these conditions. Figs. 5(a) and (b) and Figs. 6(a) and (b) show the results with the largest number of correct points in the cases with 45 and 90 images, respectively. On the other hand, in Fig.5(c), with fewer correct points, there are fewer hollows on the face, and the figure is relatively smooth, while there are some strange hollows near the goggles and the chin. Also, in the case of 90 images, as shown in Figs. 6(c) and (d), the reconstructed face at the final step is smoother and less rugged than those of earlier steps, while there is a strange precipice on the left side. The face is not well reconstructed in Figs. 5(a) and (b) and Figs. 6(a) and (b), although the numbers in Table 1 indicate better results than in the case of Figs. 4(c) and (d). This may be because the population parameter, the total number of tracked points with or without error, is not constant. Some areas tracked correctly give a large number in Table 1, even if quite rugged areas exist. For the purpose of reconstructing the human face, the total balance in the result should be more important than the high accuracies only in some areas. For example, Fig. 5(c)

Table 1. Number of points successfully reconstructed.

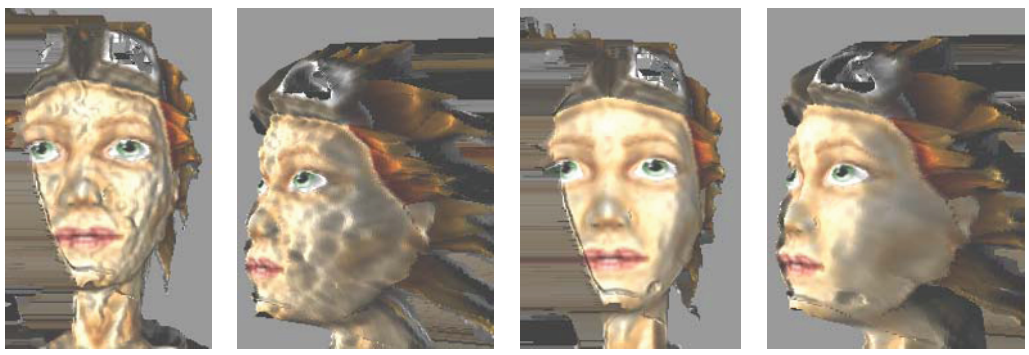
interpolation between pixels	spline	linear	spline	spline	spline	spline	spline	spline
interpolation between vectors	linear	linear	spline	linear	linear	linear	linear	linear
number of pictures	180	180	180	45	90	250	180	180
algorithm for distance calculation	1free	1free	1free	1free	1free	1free	3free	direct
size of initial search window	5	5	5	5	5	5	5	5
number of correct points after 1st calc of correspondences (1st step)	1000	874	1000	1740	1420	856	990	1030
number of correct points after 2nd calc. of correspondences (3rd step)	1460	1140	1390	1480	1790	1320	1470	1450
number of correct points in final results	1590	1420	1510	795	1730	2390	1570	1610



(a) (b) (c)

Fig. 5. Images reconstructed from 45 pictures:
(a) front view after step 1, (b) side view after step 1, (c) side view, final.

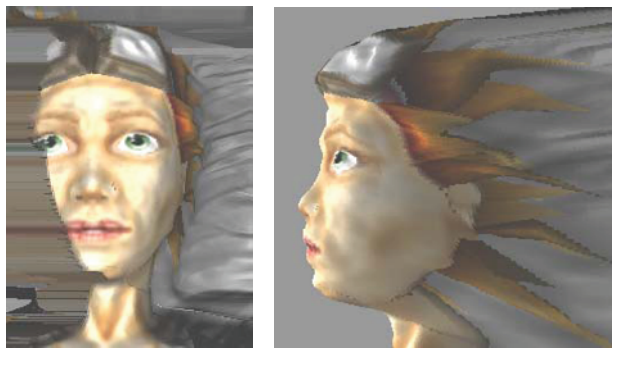
can render the characteristics of the face better than Figs. 5(a) and (b). Therefore, a more reliable method for the evaluation of this system should be developed. The result from 180 images is better than those from 45 or 90 images in our system. 3D shape measurement from fewer images needs a more robust and accurate method of searching and analyzing points for correspondences.



(a) (b) (c) (d)

Fig. 6. Images reconstructed from 90 pictures:
(a) front view after step 3, (b) side view after step 3, (c) front view, final, (d) side view, final.

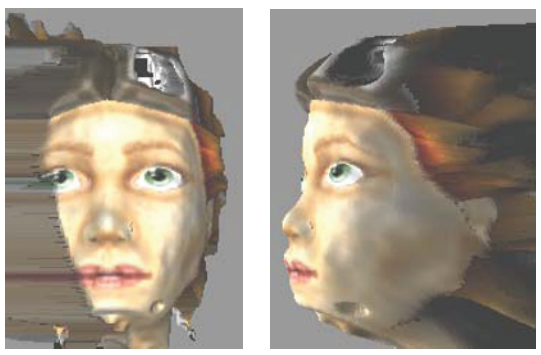
Figures 7 show the front view and side view reconstructed from 250 images. The number of correct points comes to 2390 at the final step. Less numbers at the earlier steps seem to be caused by tracking the circular streaks. Even if erroneous correspondences are not so great, errors from tracking points are integrated. However, applying the predicted vectors modifies the integrated errors and refines the result at least in these examples. The 3D shape is well reconstructed overall, i.e., near the goggle, hair, chin, nose, forehead and so on. This result, like those of Figs. 4(c) and (d), is good enough for representing a human face. Therefore, it is better to analyze many images, 180-250 pictures, in our system, even though analyzing many images requires much time and may induce larger integrated errors. It is very important to optimize the number of images and the system itself. Figures 8 and 9 show the reconstructed images by the three-freedom method and by the direct method, respectively. The numbers of points with almost correct distance increase gradually up to 1570 and 1610, respectively. These numbers are not so different from those in the case of the one-freedom method at least in analyzing the synthetic images of TINY. Each algorithm to estimate the distance of points has following characteristics. The one-freedom method is robust for the global occlusions due to its reciprocal tracking. It does not smooth the reconstruction so much. The



(a) (b)

Fig. 7. Images reconstructed from 250 pictures ; (a) front view, final, (b) side view, final.

three-freedom method may be much more robust for global occlusions due to the greater number of degrees of freedom, although it can not help to smooth the reconstruction, as shown in Fig. 9. There is a tradeoff between robustness and sharpness. The number of correct points is the largest in the case with the direct method, at least in this simulation. Figure 8 show the face of TINY well, but is a little rough, and may be too sensitive with respect to occlusions and other noises.



(a) (b)

Fig. 8. Images reconstructed using the three-freedom method;
(a) front view, final, (b) side view, final.

(a) (b)

Fig. 9. Images reconstructed by the direct method;
(a) front view, final, (b) side view, final.

4. Conclusion

A new and simple system for 3D shape measurement, which is composed by a single camera and a rotating periscope, is presented. Many pairs of stereo images are captured at intervals of small angle of periscope rotation. The distance along the optical axis of camera from the focal point of the camera system to a certain point on the object surface is inversely proportional to the disparity. In this system, a circular streak is obtained by tracing a point on each image with the cross-correlation method in the direction of periscope shift. A distance can be calculated from the radius of streak.

Applying this method to shape measurement of a synthetic human face, it is well reconstructed. This system may be adaptable to a simple handy 3D shape acquisition system. While a more reliable method for the evaluation of this system should be developed, we also examine the capability of this system using three kinds of algorithms for estimating the shape and several conditions of the different number of used images. In order to avoid incorrect distances caused by the tracking process for circular streaks, it is an important task for future studies to find some way to avoid integrated errors. In this paper, we only calculate the distances for points in the first image only. Another step of the algorithm must be developed in order to get distances for points in multi-viewed images.

Acknowledgments

The present study was supported by a grant under the Industrial Technology Research Grant Program (2000) from the New Energy and Industrial Technology Development Organization (NEDO) of Japan. The authors would also like to acknowledge the helpful technical support of Mr. M. Inaba.

References

Okutomi, M. and Kanade, T., A Multiple-Baseline Stereo", IEEE. Trans. PAMI., 15-4(1993), 353-363.

Zheng, J. Y. and Murata, A., Acquiring a Complete 3D Model from Specular Motion under the illumination of Circular-Shaped Light Sources, IEEE. Trans. PAMI., 22-8(2000), 913-920.

Zitnick, C. L. and Kanade, T., A Cooperative Algorithm for Stereo Matching and Occlusion Detection, IEEE. Trans. PAMI., 22-7(2000), 675-684.

Author profile

Satoshi Someya: After receiving his Ph.D. in Nuclear Engineering in 1998 from the University of Tokyo, he worked as a Research Fellow of New Energy and Industrial Technology Development Organization (NEDO) (1998-2000) in the Mechanical Engineering Laboratory of AIST. He has worked at the National Institute of Advanced Industrial Science and Technology (AIST) of Tsukuba as a temporary researcher. He is a member of the Thermal Engineering Research Group. His research interests include Crystal Growth, CO2 sequestration, Flow Induced Vibration and Flow Visualization.

Koji Okamoto: He received his MSc (Eng) and Ph.D. in Nuclear Engineering in 1985 and 1992, respectively, from the University of Tokyo. He worked in the Department of Nuclear Engineering at Texas A&M University as a visiting associate professor in 1994. He has been working in the Nuclear Engineering Research Laboratory and at the University of Tokyo as an associate professor since 1993. His research interests include Quantitative Visualization, PIV, Holographic PIV, Flow Induced Vibration and Thermal-hydraulics in the Nuclear Power Plant.

Gentaro Tanaka: He received his MSc (Eng) in Nuclear Engineering in 1999 from University of Tokyo. He worked in the Nuclear Engineering Research Laboratory at the University of Tokyo as a Research Associate from 1999 to 2002. He has been working in Bain & Company as a Associate Consultant since 2002. His research interests include laser specklegram, free surface turbulence, PIV, and stereoscopic PIV.