Integration of Multiple-baseline Color Stereo Vision with Focus and Defocus Analysis for 3D Shape Measurement

Ta Yuan and Murali Subbarao
tyuan@sbee.sunysb.edu and murali@sbee.sunysb.edu
Department of Electrical Engineering, State University of New York
Stony Brook, NY 11794-2350, USA

ABSTRACT

A 3D vision system named SVIS is developed for three-dimensional shape measurement that integrates three methods: (i) multiple-baseline, multiple-resolution Stereo Image Analysis (SIA) that uses color image data, (ii) Image Defocus Analysis (IDA), and (iii) Image Focus Analysis (IFA). IDA and IFA are less accurate than stereo but they do not suffer from the correspondence problem associated with stereo. A rough 3D shape is first obtained using IDA and then IFA is used to obtain an improved estimate. The result is then used in SIA to solve the correspondence problem and obtain an accurate measurement of 3D shape. SIA is implemented using color (RGB) images recorded at multiple-baselines. Color images provide more information than monochrome images for stereo matching. Therefore matching errors are reduced and accuracy of 3D shape is improved. Further improvements are obtained through multiple-baseline stereo analysis. First short baseline images are analyzed to obtain an initial estimate of 3D shape. In this step, stereo matching errors are low and computation is fast since a shorter baseline results in lower disparities. The initial estimate of 3D shape is used to match longer baseline stereo images. This yields more accurate estimation of 3D shape. The stereo matching step is implemented using a multiple-resolution matching approach to reduce computation. First lower resolution images are matched and the results are used in matching higher resolution images. This paper presents the algorithms and the experimental results of 3D shape measurement on SVIS for several objects. These results suggest a practical vision system for 3D shape measurement.

Keywords: Stereo vision, multiple-baseline stereo, color stereo, focus analysis, defocus analysis, 3D shape measurement

1. INTRODUCTION

Three-dimensional (3D) shape measurement techniques in machine vision have varying requirements in terms of amount of image data, computational resources, and camera hardware. The techniques yield varying levels of performance in terms of accuracy and speed. The performance of many of these techniques can be improved if an initial rough estimate of shape is available in addition to the required image data and camera parameters. The initial rough estimate of shape can be obtained using another technique with lesser requirements (e.g. less computation or image data) . This suggests integrating two or more compatible techniques to optimize the overall performance. In our previous paper[11] we presented a method for three such techniques — Image Defocus Analysis (IDA),[8,14,5,4] Image Focus Analysis (IFA)[3,13,10,9], and Stereo Image Analysis (SIA).[2,12,15,6,7,11] In this paper we enhance the previous method in three respects — (i) using color (RGB) image data instead of monochrome image data in Stereo Image Analysis, (ii) using stereo images recorded at multiple-baselines, and (iii) using a multiple-resolution stereo matching algorithm.

In IFA, a large sequence of image frames of a 3D scene is recorded with different camera parameters (e.g. focal length or/and lens to image detector distance). In each image frame, different objects in the scene will be blurred by different degrees depending on their distance from the camera lens. Each object will be in best focus in only one image frame in the image sequence. The entire image sequence is processed to find the best focused image of each object in the 3D scene. The distance of each object in the scene is then found from the camera parameters that correspond to the image frame that contains the best focused image of the object.

In IDA only a few (2 or 3) image frames of a scene are recorded with different camera parameters. The degree of defocus of each object in these few image frames along with the corresponding camera parameters are analyzed to find the focused image and distance of every object in the scene. In comparison with IFA, IDA requires (i) less number of images, (ii) less computation, but (iii) more information about the defocusing characteristics of the camera. In addition, IDA is less accurate than IFA. Integration of the two techniques will obtain accuracy equivalent to IFA but with less number of images

and computation than IFA. This is accomplished by first using IDA to obtain a rough estimate of depth-map, and improving the accuracy of the estimate using IFA in a narrow range around the estimated depth-map.[11]

In SIA two or more images are recorded from two or more spatial locations by displacing the camera. Then distance of objects in the scene are found through "triangulation". Depending on camera parameter values, in some vision systems (e.g. the human vision system), IDA and IFA methods are less accurate than stereo vision in providing the depth-map of a scene. However, unlike stereo vision, IDA and IFA do not suffer from the *correspondence* and *occlusion* problems. Our previous work[11] reported a technique for integrating IFA and IDA with stereo vision. The rough depth-map provided by IDA and IFA was used to simplify the stereo correspondence and occlusion detection problems. The rough depth-map essentially reduces the range of stereo disparity which is searched for stereo matching. In addition, false matches due to occlusion are reduced. Therefore, stereo analysis yields a more accurate 3D shape of objects. If the object whose 3D shape is to be measured does not have sufficient contrast information, then contrast is introduced by projecting a light pattern onto the object. This facilitates the application of IFA, IDA, and SIA.

Color images provide more information than gray-level images. Therefore, using color images in stereo matching yield more accurate estimates of disparity than gray-level images (this can be shown by constructing counter examples of color objects). But color image matching requires about 3 times more computation.

In the multiple-baseline SIA, three or more images are recorded at positions along the same baseline. Longer baseline yields better precision due to wide triangulation. It also increases the possibility of false matches due to larger range of search for the best match along the epipolar line. Shorter baseline suffers less precision but has the advantage of smaller range of search for the best match. It also reduces computational time in stereo matching. The matching results obtained using shorter baseline images can be used to estimate approximate disparity for longer baseline images. This estimate can be used to match longer baseline images with less computation and also less error.

The computations in stereo matching could be reduced by first matching images at low spatial resolution and using the results of this step to match high resolution images. In this paper, we implement multiple-resolution matching approach in the single-baseline SIA to show its effectiveness. The overall 3D shape measurement technique integrates IDA, IFA, and SIA. It is implemented on a camera system named Stonybrook VIsion System (SVIS) (see Fig. 1). Algorithms and results of experiments on SVIS are presented.

2. COLOR STEREO MATCHING (COLOR DEFOST)

The 3D shape measurement algorithm presented in our previous paper[11] is named DEFOST and it integrates Defocus analysis (IDA), Focus Analysis (IFA), and Stereo Analysis (SIA). Color image data could be used easily in IDA and IFA. For example, at each pixel, the color band with the highest contrast (or highest Laplacian) is determined from among the RGB band images. Then the IDA/IFA is applied to that band to get an estimate of depth at the corresponding pixel. However color data is not used by us in IDA and IFA in this paper since the resulting improvement in performance was thought to be marginal.

In DEFOST, single-baseline stereo was implemented using gray-level images. For each small image region in the right image, the best match in the left image was found by minimizing the Sum-of-Squared-Difference (SSD [(see 6)]) measure defined by:

$$SSD = \sum_x \sum_y \left| f_r(x,y) - f_l(x,y) \right|^2 . \tag{1}$$

In the above equation, f_r, f_l are image gray-levels of the right and left images, respectively; and *(x, y)* is the index in the matching window of a pre-determined size.

In color stereo matching, we use Color SSD defined as the sum of the SSDs computed for each of the three color bands:
$$Color\ SSD = SSD_{red} + SSD_{green} + SSD_{blue} . \tag{2}$$

The matching computation for color images is 3 times that of gray-level images.

2.1. Experiments
The Stonybrook VIsion System (SVIS) (see Fig. 1) described in our previous paper[11] was used in the experiments. All parameters of the camera system were the same as those used in that paper. The DEFOST algorithm for integrating IDA,

IFA, and SIA, presented in that paper was modified to use color image data for stereo matching. IDA and IFA were applied to gray-level images computed from color images as

$$\textit{Grey-level value = (Red + Green + Blue)/3 .}\tag{3}$$

The 3D shape of a prism placed about 0.65 meter was measured using the DEFOST algorithm with both gray-level and color image data. Objects used in the experiments are shown in Fig. 13, Fig. 14, and Fig. 15. The results are shown in Fig. 2 and Fig. 3 for comparison. A color random dot pattern was pasted on the prism to create a high contrast image. Instead of pasting a pattern, high contrast images can also be created by projecting a light pattern with color random dots. For the prism object used in our experiments, the improvement in the accuracy of 3D shape obtained using color image data instead of gray-level image data is small.

3.   MULTIPLE-RESOLUTION SIA

Computation time in stereo image matching can be reduced by a coarse-to-fine search strategy. First a coarse search is made for the best match under low spatial resolution. Lower resolution images are obtained by down sampling the original images. Spatial resolution is typically reduced by a factor of 2 to 4. The disparity-map obtained under low-resolution matching is scaled to original high resolution and used as initial estimates for high resolution matching. In high resolution matching, the search for the best match is limited to a small space along the epipolar line around the initial estimate. The size of the search space is determined by the resolution scaling factor and the expected error in the low resolution matching.

Multiple-resolution matching technique could be applied to color image data by multiple-resolution processing of each band image. However, in this paper, we apply multiple-resolution matching to only gray-level images. Multiple-resolution image analysis may also be applied to IFA, but it is not considered in this paper.

3.1.   Experiments

Multiple-resolution matching algorithm was implemented as follows. First, the IDA and IFA part of the DEFOST algorithm was implemented using the camera parameters and the algorithm as described in our previous paper.[11] Only the image size was chosen to be 448×448 instead of 432×432 to make the lower resolution image size be a multiple of 16. After applying IDA and IFA, we obtain a 28×28 depth-map where each depth estimate corresponds to one 16×16 image block. The depth values are expressed in terms of the focusing step number of the stepper motor that moves the camera lens forward and backward. In addition to depth-map, IDA and IFA also yield left and right focused images of size 448×448. These images are shrunk to 224×224 by down sampling the original images by a factor of 2. The depth-map was averaged (in units of focusing step number) in 2×2 blocks to obtain a 14×14 low resolution depth-map. The maximum error in the depth-map was taken to be ±6 steps as in DEFOST. This maximum error was used to look up a calibration table to determine the range of disparities over which a search for best match should be made. This calibration table was obtained by scaling the corresponding table used in our previous work. Scaling of the table by a factor of 2 accounts for image shrinking by the same factor.

Stereo matching of the 224×224 low resolution left and right focused images was carried out using 16×16 image blocks and *SSD* as a measure of similarity. The range of search for the best match was determined from disparity range as mentioned above. The result of this step was a 14×14 depth-map array (expressed in pixels of disparity) with improved accuracy. This depth-map was magnified to a 28×28 array by expanding each depth estimate to a 2×2 array and the disparities were scaled by 2. The resulting 28×28 depth-map array was taken to be the initial disparity estimate for the 448×448 high resolution (original) images. The maximum error in this disparity estimate was taken to be ±10 (pixels). These estimates were refined through stereo matching on 16×16 image blocks on the high resolution images using *SSD* as before. The resulting disparities were used to obtain actual depth estimates (expressed in mm) using a calibration table as in our previous work.

The results thus obtained (multiple-resolution matching) compare well with the ones obtained using the original DEFOST (only high-resolution matching). Further, as expected, the multiple-resolution approach was found to be faster by a factor of nearly 2. The results for two objects — a prism and a face — are presented (see Fig. 4, Fig. 5, Fig. 6, and Fig. 7; Fig. 5 is the low-resolution result and Fig. 7 is the result of the DEFOST.) The computation times are tabulated below.

Stereo matching	Multiple-resolution		High resolution only
Resolution	Low resolution	High resolution	
Computation time (sec)	6	16	40

Total time (sec)	6+16=22	40

Table 1 Comparison of computation time for multiple-resolution and single resolution stereo matching

The speedup is almost 2 (40/(6+16)) in our case. In the multiple-resolution matching approach, lower the resolution, higher the matching uncertainty and therefore larger the search space at the higher resolution. Also, if the resolution is too low, gross matching errors are possible. Accurate estimation of the matching uncertainty is also difficult.

4.   MULTIPLE-BASELINE SIA

The two problems associated with multiple-baseline stereo are accuracy and precision.[6,7]   Longer baseline yields better precision due to wider triangulation. But longer baseline makes the range of search for the best match larger and leads to a higher possibility of false matches. In [6,7] M. Okutomi, et al, propose that using multiple-baseline stereo can determine a unique and clear minimum of sum of sum of squared differences (*SSSD*)-in-inverse-depth at the correct matching position and can also eliminate false matches and increase precision. But they have to search the whole possible depth range to compute the *SSSD*. In our case we have a good initial estimate of the depth map from IDA/IFA. This estimate is used to limit the search range and reduce false matches. In the best case the unique minimum of *SSD* will be found in a range around the estimated depth map[11] obtained by IDA/IFA.

4.1.   Experiments

In DEFOST we take two images, say, right and left, and in our multiple-baseline configuration we take a third image (the middle one) and form two pairs of stereo images. The right and the middle ones form the first pair and the right and the left form the other. We first apply SIA on the first pair of stereo images. It refines the result of the IDA/IFA output. Because of the shorter baseline of the first pair the precision is less than the second pair but the search range in matching will also be smaller. This yields a less precise result but faster matching. This result will be used to obtain an initial estimate of the disparity map for matching the second longer-baseline pair of stereo images. The second pair of images are matched by searching in a narrow disparity range around the initial estimates. Thus the computation is reduced and the accuracy of the final matching results are improved.

The results of 3D shape recovery by multiple-baseline stereo using gray-level image data is presented for a cone object here (see Fig. 8). This result compares well with the result of the original DEFOST. All camera parameters for both methods are the same. The computation time of multiple-baseline stereo matching for several objects are averaged and tabulated below (see Table 2) for several baseline factors (the factor by which the full baseline is divided). For full baseline (50 mm) the baseline factor is 1 and for half baseline (25mm) the factor is 2. The disparity vs depth calibration data for different baselines were obtained by scaling the calibration data for full baseline with the baseline factor. The last row of the table lists the computation speed-ups obtained.

Baseline factor	8	4	2	4/3	1
Time for matching the first pair	6	14	24	34	40
Time for matching the second pair	7	7	7	7	N/A
Total matching time	13	21	31	41	40
Speedup in time	3.08	1.90	1.29	0.98	1

Table 2 Speedup in time for multiple-baseline SIA

In the case of 1/8 baseline the speedup is high but the final result is less accurate than the rest because the precision error may overshadow the accuracy in the second pair matching. If we increase the search range for the second pair then the computational time will also increase. There is a tradeoff between speed and accuracy. On the other hand obtaining the third focused image takes time. Our camera system takes about 2 seconds to acquire each image. If the depth values of adjacent image blocks have large difference then there will be a large amount of distortion in a stereo image pair due to foreshortening effect. The longer the baseline is, the larger the effect is. For an object of large depth variations and a long baseline, the matching will not be good due to foreshortening effect. We found that adding a third image in the middle and using the shorter-baseline stereo image pair to get an initial result will help the matching on the longer-baseline stereo image pair. This could reduce false matches and increase accuracy of matching.

4.2. Multiple-baseline SIA with color image data

We implemented the multiple-baseline SIA with color image matching. Results for two objects, a cone object and a face object, are presented for a baseline factor of 2. We also show the intermediate results of the first stereo image pairs (for half baseline) (Fig. 9 – Fig. 12). These results compare well to DEFOST and color DEFOST. We also tabulate below the computation times for various baseline factors. The times are obtained by averaging the results of several experiments.

Baseline factor	8	4	2	4/3	1
Time for matching the first pair	27	41	78	106	127
Time for matching the second pair	25	25	25	25	N/A
Total matching time	52	66	103	131	127
Speedup in time	2.44	1.92	1.23	0.97	1

Table 3 Speedup in time in color multiple-baseline SIA

5. CONCLUSION

The experimental results on SVIS indicate the feasibility and advantages of integrating multiple-baseline color stereo with IDA, IFA, and multiple-resolution matching technique. The advantages include improved accuracy of 3D shape and reduction in computation time. This work can be extended to use more than two baselines and also to apply multiple-resolution matching technique to multiple-baseline stereo with color image data. It can also be extended to use other stereo matching techniques instead of the *SSD* based method.

6. ACKNOWLEDGEMENTS

7. REFERENCES

1. B. Ross, " A Practical Stereo Vision System", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York City, New York, pp. 148-153, June 1993.
2. B.K.P. Horn, *Robot Vision*, McGraw-Hill Book Company, 1986.
3. E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, 223-237, 1987.
4. G. Surya, "Three-dimensional scene recovery from image defocus", Ph.D. Thesis, Dept. of Electrical Eng., SUNY at Stony Brook, 1994.
5. J. Enns and P. Lawrence, "A Matrix Based Method for Determining Depth from Focus", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1991.
6. M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo", *IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, 1991.
7. M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo", *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 15, no. 4, April 1993.
8. M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach", *International Journal of Computer Vision*, 13, 3, pp. 271-294 (1994).
9. M. Subbarao, and J.K. Tyan, "The Optimal Focus Measure for Passive Autofocusing and Depth-from-Focus", *Proceedings of SPIE conference on Videometrics IV*, Philadelphia, Oct 1995.
10. M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *Journal of Optical Engineering*, Vol. 32 No. 11, pp. 2824-2836, November 1993.
11. M. Subbarao, T. Yuan, J.K. Tyan, "Integration of Defocus and Focus Analysis with Stereo for 3D Shape Recovery", *Proceedings of SPIE Conference on Three-Dimensional Imaging and Laser-Based Systems for Metrology and Inspection III*, Vol. 3204, Pittsburgh PA, October 1997
12. Robert M. Haralick and Linda G. Shapiro, *Computer and Robot Vision*, Vol. II, Addison-Wesley Publishing Company, 1993.
13. S. K Nayar, "Shape from Focus System", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 302-308 (June 1992).
14. T. Wei, "Three-dimensional machine vision using image defocus", Ph.D. Thesis, Dept. of Electrical Eng., SUNY at Stony Brook, 1994.
15. U.R. Dhond, J.K. Aggarwal, "Structure from Stereo - A Review", *IEEE Trans. on System, Man, and Cybernetics*, vol. 19, no. 6, November/December, 1989.
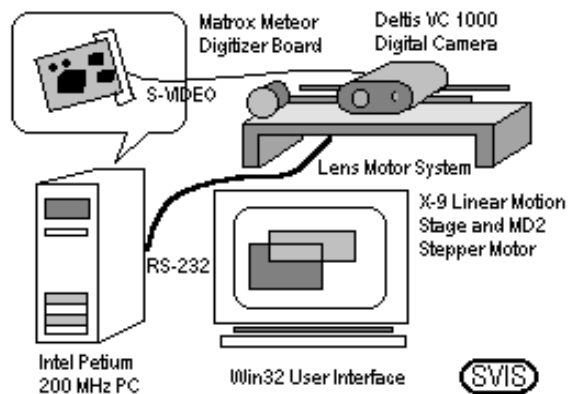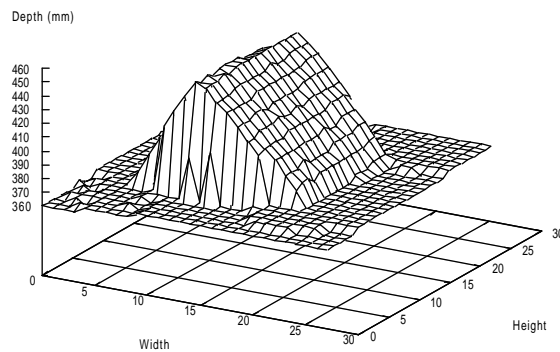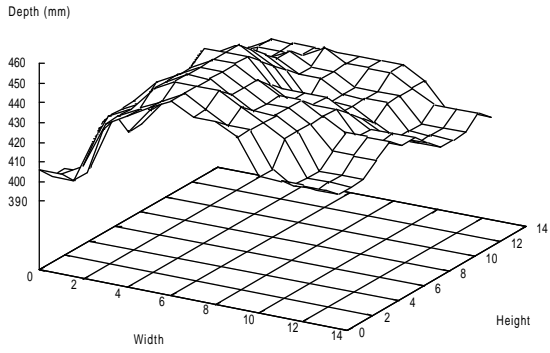
Fig. 1 SVIS Camera System
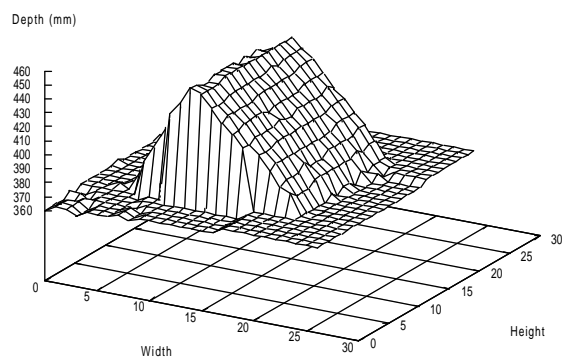
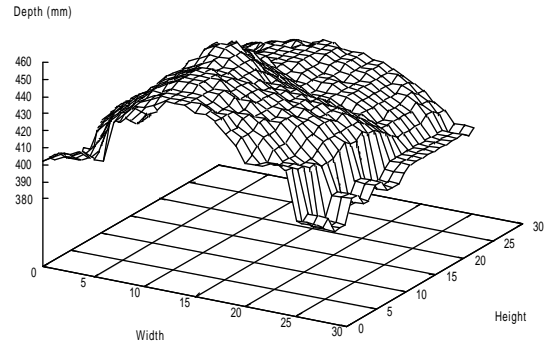**Fig. 2 Result of DEFOST for the prism object**
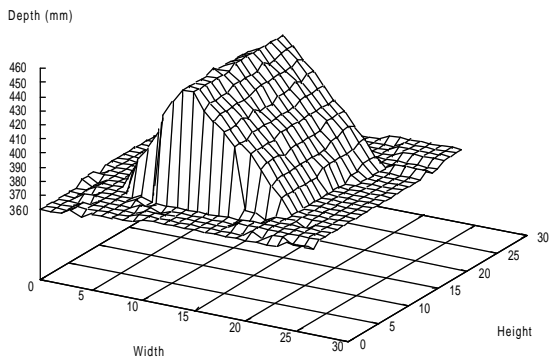
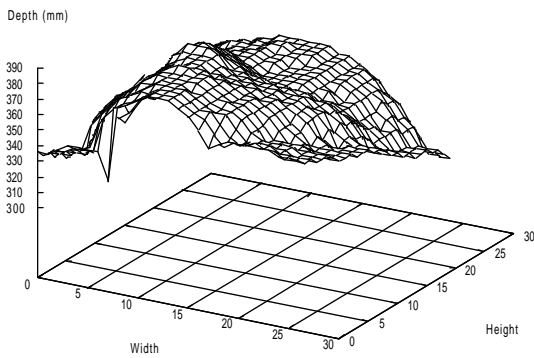**Fig. 3 Result of Color DEFOST for the color prism object**

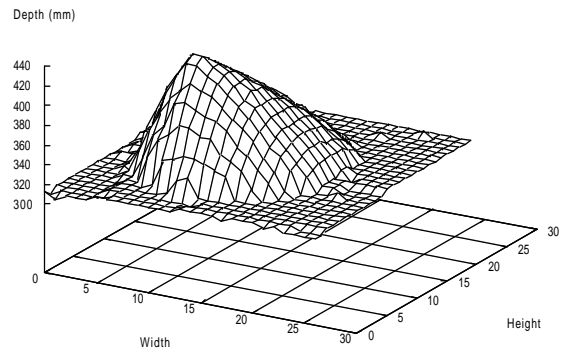**Fig. 4 Result of multiple-resolution SIA for the prism object**

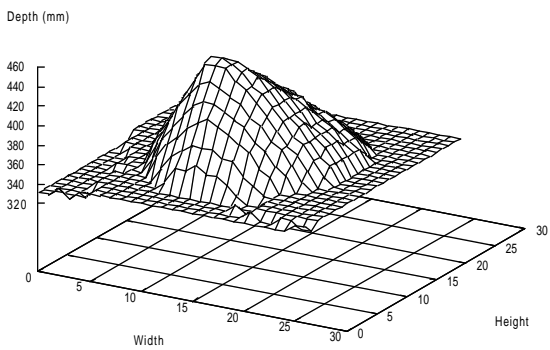**Fig. 5 Low resolution result of multiple-resolution SIA for the face object**

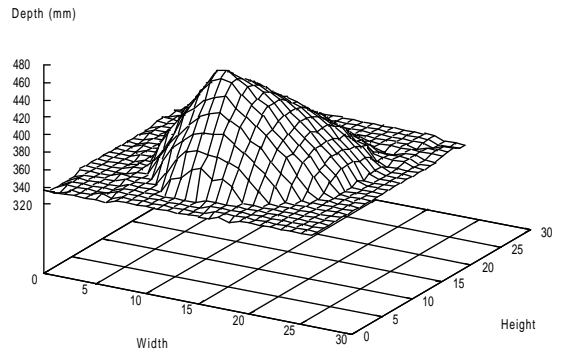**Fig. 6 Result of multiple-resolution SIA for the face object**

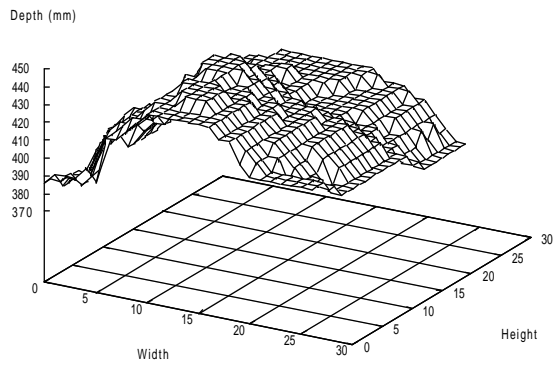**Fig. 7 Result of DEFOST for the face object**

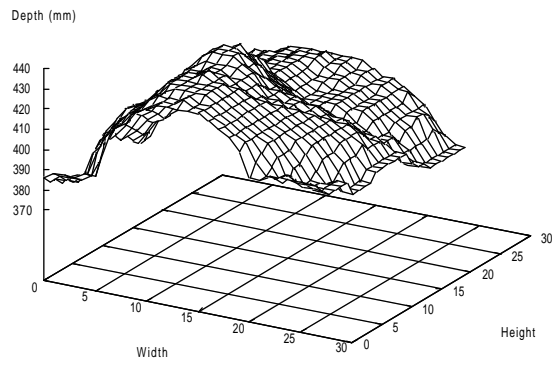**Fig. 8 Result of multiple-baseline SIA for the cone object**

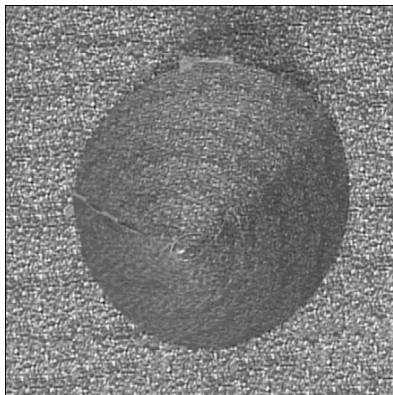**Fig. 9 Result of the first stereo pair of color multiple-baseline SIA for the color cone object**

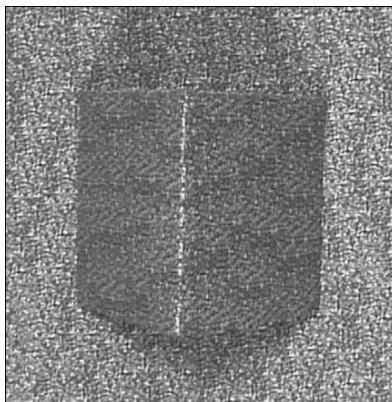**Fig. 10 Result of color multiple-baseline stereo for the color cone object**

Fig. 11 Result of the first stereo pair of color multiple-baseline SIA for the color face object

Fig. 12 Result of color multiple-baseline SIA for the color face object

Fig. 13 Picture of the cone object

Fig. 14 Picture of the prism object

Fig. 15 Picture of the face object