A Dynamic Vision System for Tracking and Localization of Underwater Objects

Sheng-Wen Cheng *, Huai-Wen Hsu ** and Shiao-Chung Cheng**
*Professor, **Graduate Student
Department of Engineering Science and Ocean Engineering
National Taiwan University, Taipei, Taiwan 106
(phone: +886-2-33665760; fax: +886-2-23929885; e-mail: niccheng@ntu.edu.tw).

ABSTRACT—**We describe the development of an underwater cross-range high-accuracy 3D metrology by utilizing a pair of video cameras for underwater vehicle applications. The novelty of the approach lies in (1) a polynomial representation of image non-linear distortion induced by lens system and surrounding water; (2) the transfer of photogrammetric bundle adjustment techniques to underwater ranging; and (3) a new representation of underwater stereo vision including camera self-calibration, match points searching, and computation of the target coordinates.**

We demonstrate that our method achieves better flexibility and accuracy compared to traditional stereographic methods.

Index Terms—**underwater vehicle, stereo vision, dynamic vision, photogrammetry, bundle adjustment, camera calibration..**

I. INTRODUCTION

Stereo vision is a common and popular method for underwater optical ranging.[1-5] In general practice, one assumes that the two optical axes of cameras are parallel to each other. Base on this assumption, we have developed a dynamic stereo vision system with compensation of image distortion and optical flow analysis.[6, 7] However, we found tuning of the camera poison and orientation, to ensure the parallelism, was much more difficult than performing stereo vision itself.

To solve this problem, the most relevant technique is the bundle adjustment [8-9], which is a photogrammetric technique to combine multiple images of the same scene into an accurate 3D reconstruction. Then bundle adjustment applies an iterative algorithm to compute optimal values for the 3D reconstruction of the scene, camera positions and orientations, by using a least-squares algorithm.

This paper presents an underwater dynamic vision system for underwater vehicles to perform tasks of observation, searching, ranging and tracking of underwater objects.

The vision system is composed of three modules for camera calibration, stereo imaging and video image processing.

In the camera calibration module, a polynomial model is proposed to compensate for the image distortion induced by lens and surrounding water. For the stereo imaging module, an epipolar line based algorithm is adopted for matching points searching. Object ranging is carried out using a bundle system model, originated in photogrammetry community, in which parameters of camera calibration and stereo imaging are coupled. The video image processing module is responsible for object movement estimation through optical flow analysis. The optical flow can be dissolved into translation, rotation and distance change components of objects.

Two experiments are conducted in this paper: 1. Localization and tracking of a moving object under diffuse illumination, and 2. 3D mapping of underwater surfaces by using a scanning point beam. As a result, the dynamic vision system demonstrates feasibility for AUV applications.

II. CAMERA CALIBRATION

A. Compensation of Image Distortion

Figure 1 shows an example image of a calibration board on which dots are equally spaced. We can identify image distortion through displaced dots. Source of mage distortion consists of two major parts: lens and surround media. Effect of surround media is depicted as figure 2. This figure demonstrates the fact that

0-7803-8541-1/04/$20.00 ©2004 IEEE.

underwater is far-sighted.

Figure 1 Example of image distortion: original image.

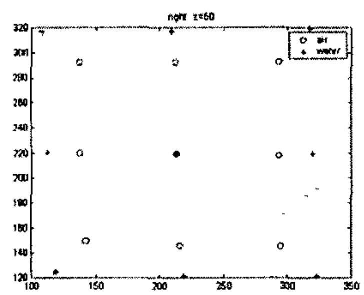


Figure 2 Effect of surrounding medis on image distortion

To compensate for the image distortion, we adopt the dual plane model proposed by Gremban [11].

$$x=\sum_{m=0}^{M}\sum_{n=0}^{N}a_{mn}x_i^m y_i^n$$
$$y=\sum_{m=0}^{M}\sum_{n=0}^{N}b_{mn}x_i^m y_i^n \quad (1)$$

where (x, y, z) is coordinates of a arbitrary point target, (x_i, y_i) is coordinates of the corresponding image point, M and N are the order of polynomial for x_i and y_i, respectively. In this paper, we select M=N=3.

Owing to the perspective effect, calibration coefficients a_{mn} and b_{mn} must be functions of z/f. Here, we use polynomial forms.

$$a_{mn}=\sum_{k=0}^{K}c_{ak}(z/f)^k$$
$$b_{mn}=\sum_{k=0}^{K}c_{bk}(z/f)^k \quad (2)$$

where K=2 in this paper.

Figure 3 Images of calibration board at z=203 and z=135 mm.

By taking images of calibration board from more than three different distances, as shown in figure 3, non-linear algebraic equations will be derived from equation (1) and equation (2). Then solving the equations by using least-square algorithm, we obtain calibration coefficients for right and left cameras as shown in table 1 and table 2, in case of underwater and dry conditions.

Table 1 Calibration Coefficients: c_R and c_L in water

	C_R		
	c_0^R	c_1^R	c_2^R
a_{00}	0.0007	-0.4602	5.8875
a_{10}	0.0012	0.9016	13.0059
a_{01}	-0.0008	0.1185	-6.5752
b_{00}	-0.0018	0.0272	-11.7107
b_{10}	0.0015	-0.1910	5.6253
b_{01}	0.0010	1.0548	9.0964
	C_L		
	c_0^L	c_1^L	c_2^L
a_{00}	0.0006	-0.3626	4.4153
a_{10}	0.0005	1.0132	5.8813
a_{01}	-0.0010	0.1229	-3.0734
b_{00}	-0.0018	0.0236	-11.0636
b_{10}	0.0010	-0.1242	4.7871
b_{01}	0.0010	1.0635	7.1424

Table 2 Calibration Coefficients: c_R and c_L in air

	C_R		
	c_0^R	c_1^R	c_2^R
a_{00}	0.0023	-0.7706	8.8642
a_{10}	-0.0085	2.5059	-23.2657
a_{01}	0.0002	-0.0208	-0.6509
b_{00}	0.0014	-0.4257	0.0880
b_{10}	-0.0004	0.0370	-1.7612
b_{01}	-0.0068	2.3952	-16.6810

	C_L		
	c_0^L	c_1^L	c_2^L
a_{00}	0.0024	-0.6478	9.2164
a_{10}	-0.0075	2.3771	-22.2621
a_{01}	-0.0010	0.1308	-2.9388
b_{00}	0.0007	-0.3473	-1.7444
b_{10}	-0.0017	0.2112	-5.7911
b_{01}	-0.0020	1.8347	-2.5880

Although calibration coefficients for underwater and dry conditions are different, the image of control points on calibration board after compensation will be almost the same as shown in figure 4.

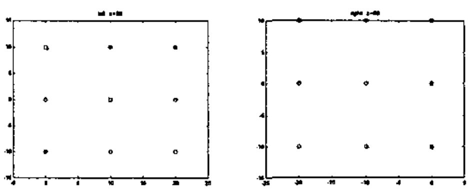


Figure 4 Left and right images of control points after compensation

B. Camera Position and Orientation (figure 5)

Considering a target point P whose coordinates are (X_w, Y_w, Z_w) and (X_c, Y_c, Z_c) in the world coordinates system and camera coordinates system, respectively. If the coordinates of camera center in the world coordinates system is (X_o, Y_o, Z_o), then

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w - X_0 \\ Y_w - Y_0 \\ Z_w - Z_0 \end{bmatrix} = \begin{bmatrix} r_{11}(X_w - X_0) + r_{12}(Y_w - Y_0) + r_{13}(Z_w - Z_0) \\ r_{21}(X_w - X_0) + r_{22}(Y_w - Y_0) + r_{23}(Z_w - Z_0) \\ r_{31}(X_w - X_0) + r_{32}(Y_w - Y_0) + r_{33}(Z_w - Z_0) \end{bmatrix} \quad (3)$$

where

$$R(\alpha,\beta,\gamma) = R_z(\gamma)R_y(\beta)R_X(\alpha) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} =$$

$$\begin{bmatrix} \cos\beta\cos\gamma & \sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma & -\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ -\cos\beta\sin\gamma & -\sin\alpha\sin\beta\sin\alpha + \cos\alpha\cos\gamma & \cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix}, \quad (4)$$

(α, β, γ) are Euler angles of camera coordinates system rotating with respect to the world coordinates system.

On the other hand, from equation (1) and equation (2), we obtain the following relation.

$$\begin{bmatrix} X_c & Y_c \end{bmatrix}^T = A \begin{bmatrix} (\frac{Z_C}{f})^2 & \frac{Z_C}{f} & 1 \end{bmatrix}^T \quad (5)$$

where $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$ (6)

$$= \begin{bmatrix} \sum c_{mn,a0} x'^m y'^n & \sum c_{mn,a1} x'^m y'^n & \sum c_{mn,a2} x'^m y'^n \\ \sum c_{mn,b0} x'^m y'^n & \sum c_{mn,b1} x'^m y'^n & \sum c_{mn,b2} x'^m y'^n \end{bmatrix}$$

is called the camera calibration matrix.

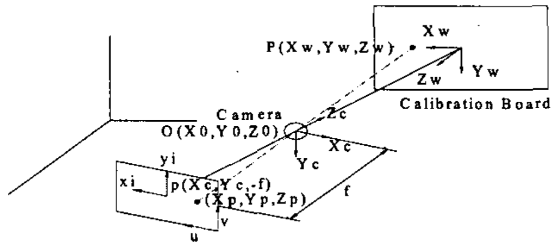


Figure 5 Calculation of Camera position and Orientation

Equation (3) and equation (5) are fundamentals of bundle adjustment in photoprammetry.

Based on Taylor's expansion, we obtain

$$\begin{bmatrix} b_{11} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ & & \vdots & & & \\ b_{n1} & b_{n2} & b_{n3} & b_{n4} & b_{n5} & b_{n6} \end{bmatrix} \begin{bmatrix} d\alpha \\ d\beta \\ d\gamma \\ dX_{w0} \\ dY_{w0} \\ dZ_{w0} \end{bmatrix} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{bmatrix} \quad (7)$$

where (X_{w0}, Y_{w0}, Z_{w0}) denotes the position of camera center in the world coordinates system.

By using iterative least square algorithm, the camera position (X_{w0}, Y_{w0}, Z_{w0}) and orientation (α, β, γ) can be solved from equation (7).

For verification, a rotation with respect to y-axis (β) test in five degree increment was conducted. Results are shown as table 3. The accuracy of the proposed algorithm is acceptable.

Table 3 Camera orientations

α (degree)	β (degree)	γ (degree)	X_{w0} (mm)	Y_{w0} (mm)	Z_{w0} (mm)
4.69	12.75	-4.27	6.37	4.51	241.76
4.80	8.53	-4.42	-14.21	1.22	239.71
4.79	3.54	-3.74	-24.06	9.21	237.90
4.93	0.02	-5.33	-53.62	-12.87	235.08
4.96	-4.56	-5.31	-68 21	-14.04	230.48
4.99	-9.34	-5.37	-79.61	-16.09	225.43

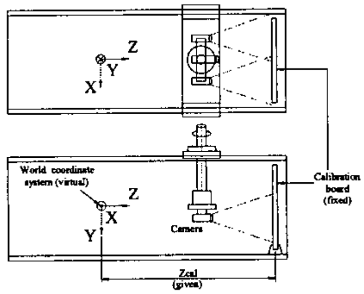


Figure 6 Configuration for camera calibration

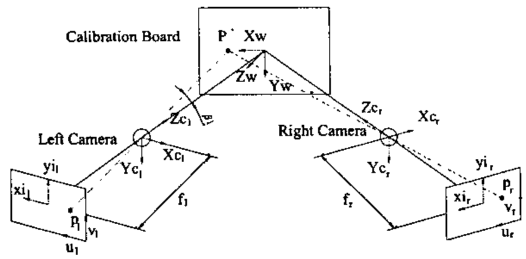


Figure 7 General Arrangement of Binocular Visual System

III. STEREO VISION USING CAMERAS OF ARBITRARY ORIENTATION

Following equation (3) and equation (5), the coordinates of target point P in figure7 will satisfy following equations.

$$\begin{bmatrix} X_{Rc} \\ Y_{Rc} \\ Z_{Rc} \end{bmatrix} = R_R \begin{bmatrix} X_w - X_{R0} \\ Y_w - Y_{R0} \\ Z_w - Z_{R0} \end{bmatrix} \quad (8)$$

$$\begin{bmatrix} X_{Lc} \\ Y_{Lc} \\ Z_{Lc} \end{bmatrix} = R_L \begin{bmatrix} X_w - X_{L0} \\ Y_w - Y_{L0} \\ Z_w - Z_{L0} \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} X_{Rc} & Y_{Rc} \end{bmatrix}^T = A_R \begin{bmatrix} (\frac{Z_{Cr}}{f_R})^2 & \frac{Z_{Cr}}{f_R} & 1 \end{bmatrix}^T \quad (10)$$

$$\begin{bmatrix} X_{Lc} & Y_{Lc} \end{bmatrix}^T = A_L \begin{bmatrix} (\frac{Z_{Cl}}{f_L})^2 & \frac{Z_{Cl}}{f_L} & 1 \end{bmatrix}^T \quad (11)$$

$$R_R = \begin{bmatrix} r_{R11} & r_{R12} & r_{R13} \\ r_{R21} & r_{R22} & r_{R23} \\ r_{R31} & r_{R32} & r_{R33} \end{bmatrix} \quad (12)$$

$$R_L = \begin{bmatrix} r_{L11} & r_{L12} & r_{L13} \\ r_{L21} & r_{L22} & r_{L23} \\ r_{L31} & r_{L32} & r_{L33} \end{bmatrix} \quad (13)$$

By eliminating (X_w, Y_w, Z_w) from above equations, non-linear equations in terms of unknowns (z_{Cr}, z_{Cl}) will be obtained.

IV. DYNAMIC VISION SYSTEM AND EXPERIMENTS

A. Experimental System Configuration (figure 8)

1.
2. Water tank and xyz Table： 2.4 m ×1.2 m × 0.8m.
3. Water-tight CCD cameras： 640 × 480 ◦ Pixels.
4. Binocular frame： shown as figure 9.
5. Calibration board.
6. Underwater Lamps.
7. Video frame grabber： Coreco Cobra/C6 card.
8. PC-based controller.
9. GPIB interface card,

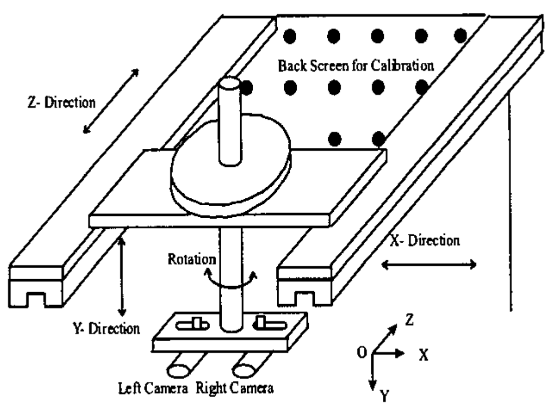


Figure 8 System Configureuration of Sensing and Perception Module Development Platform

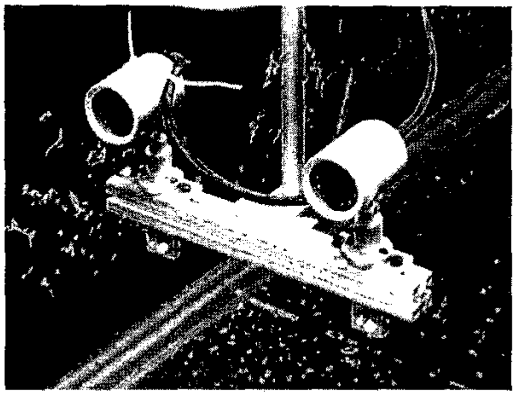


Figure 9 Binocular frame

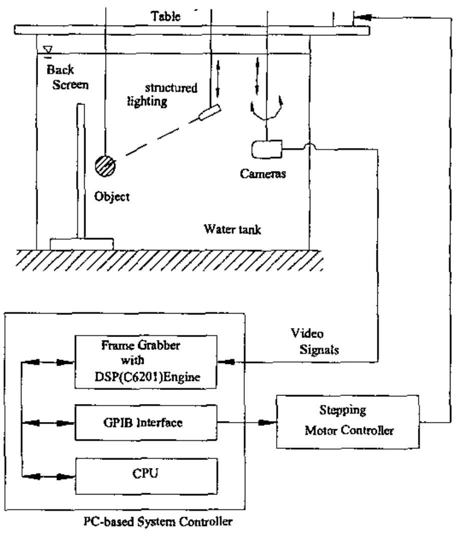


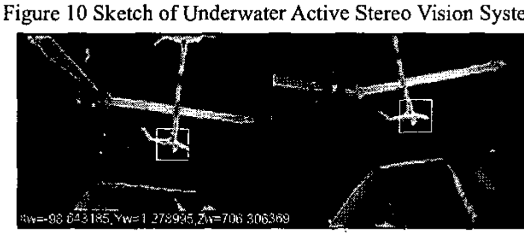
Figure 10 Sketch of Underwater Active Stereo Vision System

B. Tracking

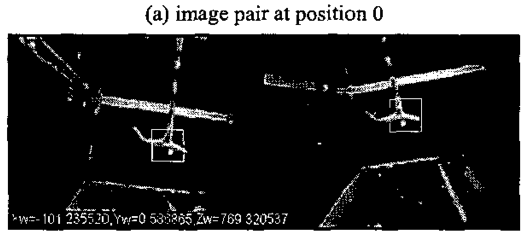
System setup is shown as figure 10. A fish-shaped object was moved along a predetermined 3D path. According to the position and speed of object, the binocular vision system adaptively followed. The degree of freedom was three: back/forth, up/down and horizontal rotation.

Image sequence is shown as figure 11. During tracking, the system identified the object based on pattern matching technique.

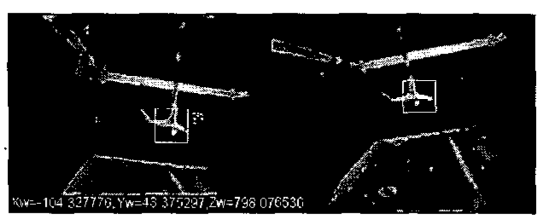
We summarize the results of object tracking and motion trajectory estimation in figure 12. The estimated path matches well with the actual one.



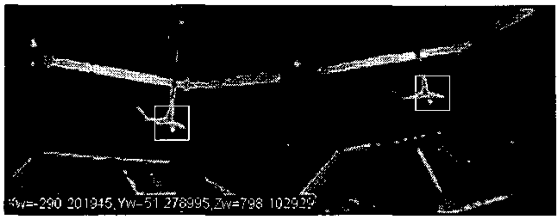
(a) image pair at position 0



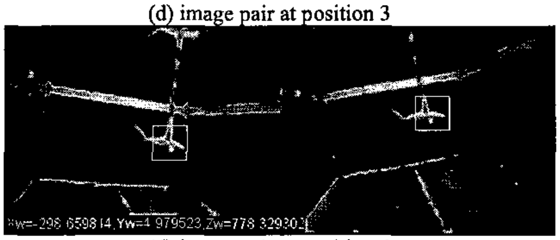
(b) image pair at position 1



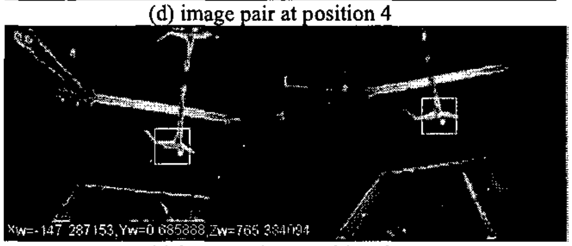
(c) image pair at position 2



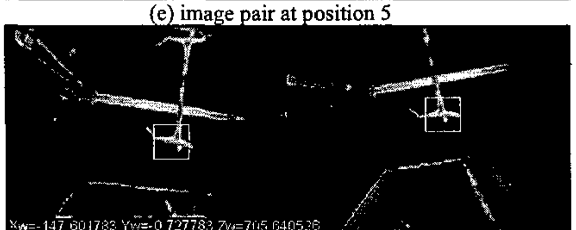
(d) image pair at position 3



(d) image pair at position 4



(e) image pair at position 5



(f)image pair at position 6

(g)image pair at position 7

Figure 11 Image Sequence

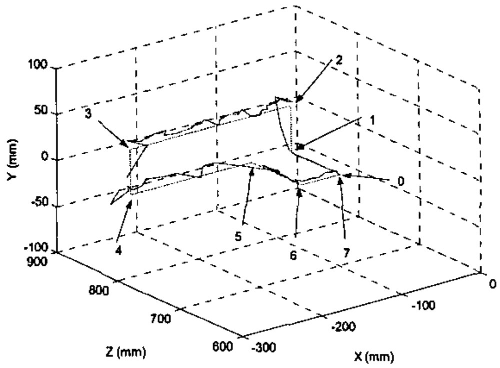


Figure 12 Object Tracking and Motion Trajectory Estimation

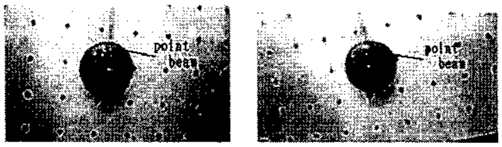
C. Mapping Using Active Laser Beam

In stereo vision, match points searching task requires high computing power. To ease hardware and software burden, we utilized active illumination method [10] and thresholding of image grayscale level. As a result, the range data acquisition nearly synchronized with frame acquisition.

Example of surface mapping is shown in figure 13. The white spots of laser beam can be easily located.

By mechanical scanning of laser point beam, we get the surface contour in 3D format shown as figure 14 and figure 15. Figure 16 and figure 17 are of 2Dcontour format.

Comparing figure14 with figure 15, and figure 16 with 17, it is clear that compensation of image distortion is essential to get accurate results.



(a) (b)

Figure 13 (a)left image; (b) right image.

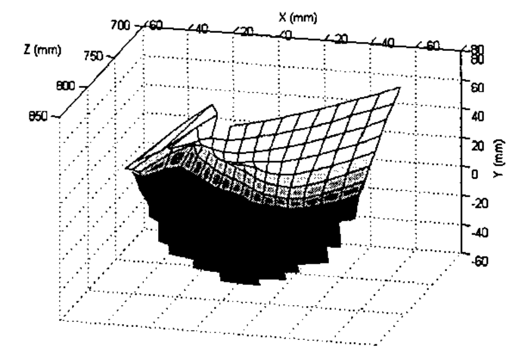


Figure 14 3D contour of underwater object (without compensation of distortion)

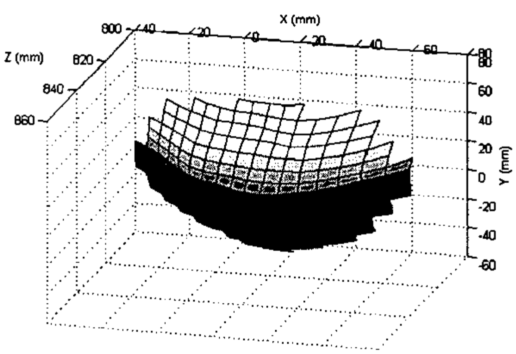


Figure 15 3D contour of underwater object (distortion compensated)

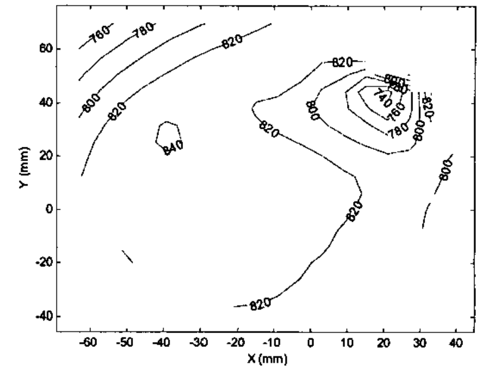


Figure 16 2D contour of underwater object (without compensation of distortion)

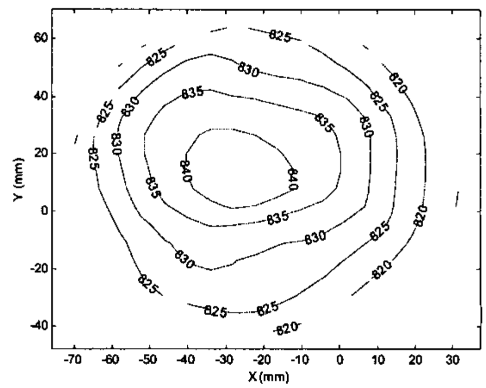


Figure 17 2D contour of underwater object (distortion compensated)

V. CONCLUSION

We demonstrated that the proposed method achieves better flexibility and accuracy compared to traditional underwater stereo vision methods.

ACKNOWLEDGMENT

The authors would like to thank the support from National Science Council of Republic of China (Taiwan).

REFERENCES

[1] Grosso E. and Tistarelli M., "Active/Dynamic Stereo Vision," Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol.17, No. 9, pp. 868 –879, 1995.

[2] Grosso E. and Tistarelli M., "Active/Dynamic Stereo: a General Framework," *Computer Vision and Pattern Recognition Proceedings CVPR '93., 1993 IEEE Computer Society Conference on,* pp. 732 -734, 1993.

[3] Inaba, M. and Hara, T., "A Stereo Viewer Based on a Single Camera with View-Control Mechanisms," Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '93, Vol. 3, pp. 1857 –1865, 1993.

[4] Rougeaux S. and Kita N., "Binocular Tracking Based on Virtual Horopters," Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems, IROS '94, Vol. 3, pp. 2052 -2057, 1994.

[5] A. Goshtasby and W. A. Gruver, "Design of a Single-Lens Stereo Camera System," Pattern Recognition, Vol. 26, No. 6, pp. 923-937, 1993.

[6] Sheng-Wen Cheng, Huai-Wen Hsu and Ling-Hao Yu, "Research on A Real-time Robotic Visual System for Ranging and Motion Estimation of Underwater Objects," Proceedings of the 4th Conference on Underwater Science and Technology (CUST'02), pp.C1-19-27, Taipei, Taiwan, April 2002.

[7] Sheng-Wen Cheng, Huai-Wen Hsu and Chun-Fu Chang, "Development of a Dynamic Vision System for Tracking and Localization of Underwater Moving Objects," Proceedings of the 5th Conference on Underwater Science and Technology (CUST'03), pp.68-75, Kao-Hsiung, Taiwan, April 2003.

[8] S.I. Granshaw, "Bundle adjustment methods in engineering photogrammetry," Photogrametric Record, Vol. 10, No, 56, pp. 181-207, 1980.

[9] K.B. Atkinson, *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, Caithness, Scotland, 1996.

[10] R. Hartley and P. Sturm, "Triangulation," Computer Vision and Image Understanding, Vol.68, No.2, pp. 146-157, 1997.

[11] Gremban, K. D. and C. E. Thorpe, "Geometric Camera Calibration Using System of Linear Equation", *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 562-567, 1988

[12] C. Stiller and J. Konrad ,"Estimating Motion in Image Sequences," IEEE Signal Processing Magazine, pp. 1053-5888, 1998.