Improving the accuracy of depth estimation using a modified stereo vision model in binocular vision

Kuan-Yu Chen[1,#], Cheng-Chin Chien[1], Wen-Lung Chang[1], and Chi-Chung Hsieh[1]

1 Department of Mechanical Engineering, Chung Yuan Christian University, Jungli City, Taoyuan County 32023, Taiwan (R.O.C.)
Corresponding Author / E-mail: gychen@cycu.edu.tw, TEL: +886-3-2654322, FAX: +886-3-2654399

KEYWORDS :Binocular vision, Stereo vision, Depth estimation, Machine vision, Image processing

Stereo vision for extraction of three-dimensional information from stereo images has been widely used in many applications like robot navigation, recovering the three-dimensional structure of a scene, and optical inspection systems. More recently, the majority of research in stereo vision has focused on the establishment of stereo matching. However, to date, there has been relatively little research conducted on the effect of computational models of binocular vision with variable focal length of lens. In this paper, a modified computational model of stereo vision is presented todevelop a new depth estimation algorithmwith no effect ofchanges in focal length. Experimental results demonstratedthat extraction of depth information from stereo pairs using the proposed method was more accurate than conventional method.

Manuscript received: JanuaryXX, 2011 / Accepted: JanuaryXX, 2011

NOMENCLATURE

B = baseline length
f = focal length of lens
I_L, I_R = horizontal positions ofanobjectintwoimages
J = vertical positions ofanobjectintwoimages
L = distance between image plane and virtual focal point
M, N = horizontal and vertical resolution of image
W, H = width and height of image
X, Y, Z = coordinates of a world point
x, y, z = coordinates of lens center

1. Introduction

Stereo or binocular vision is an active research area in the field of machine vision that attempts to imitate the distance and depth-perception abilities of the human visual system by using two cameras. In general, depth information can be used to recovery the three-dimensional structures of the shapes in a scene, track moving objects for visual navigationin automated robot control, or measure distance information for optical inspection systems[1-3].Thus, there are two major problems of a stereo vision system which are stereo matching and three-dimensional reconstruction. Over the past few decades, a number of different approaches have been proposed in the literature for solving these problems.One early research done in the field of stereo vision was by Marr and Poggio [4]. They analyzed the computational structure of the stereo-disparity problem and then proposed a cooperative algorithm to implement the computation of disparity information from two stereo images. Barnard and Fischler [5] reviewed various computational stereo techniques and provided a representative sampling of computational stereo research through 1981. Brown *et al*. [6] further reviewed advances in computational stereo roughly from the early 1990s to early 2000s, focusing primarily on three topics: correspondence techniques, methods for occlusion, and real-time implementations.However, extraction of exact depth is not necessary in many of real-time stereo vision tasks such as visual servoing, visual navigation, and obstacle avoidance. Therefore,several studieshave presented algorithms for computing the relative depth of points in a stereo image pair for active vision systems with no camera calibration or prior knowledge of the parameters of the stereo vision system required [7, 8].In addition, because of color images can provide more information than grey-level images, considerable research work has been done to offer various methods for matching color images [9].

As just mentioned, the majority of research in stereo vision has focused on the establishment of stereo matching and all of them assumed exact matching of focal lengths of the cameras.However, in recent years variable focal length lens systems have become more commonly available for cell phones, digital cameras, and machine vision applications. Sengupta [10] analyzed the effects of unequal focal lengths and derived the equations of modified epipolar lines.Furthermore, severalrecent studies have reported dealing with the design problem of operation mechanism and the effects of image acquisition with extended depth of field for variable-focus liquid lens [11, 12].We should notice that there is an important issue of stereo vision by using variable focus cameras. If the focal length of lens is not fixed or exactly known, the depth estimation error of conventional stereo imaging model will increase. To date, relatively little research has been conducted on the computational model of binocular vision with variable focus cameras.Thus, the objective of this paper is to develop a more accurate algorithmfor computing three-dimensionalinformation from a stereo pair of images by modification of the stereo vision model.

The remainder of this paper is organized as follows. Section 2 briefly reviews the fundamentals of the conventional stereo vision. Section 3 presents the proposed modified model of binocular vision. Experimental results and comparisons with conventional method are presented in Section 4. Finally, conclusions are given in Section 5.

2. The Geometry of Conventional Stereo Vision

The conventional stereo vision is to acquire a pair of images simultaneously by using two horizontally placed cameras, as shown in Fig. 1. The central problem in stereo vision, called correspondence problem, is the search for the correct match of a point in both stereo images. Here we assume that the geometry of two cameras is known and the corresponding points in two images are identified, then the geometry of stereo vision can be simplified as a triangulation. In Fig. 1, the goal of stereo vision is to find the coordinates(X, Y, Z) of the world point P having corresponding points(x_1, y_1) and (x_2, y_2) in left and right images, respectively [13]. This is easily done by the use of similar triangles, that is,

$$\frac{\frac{B}{2}-X}{x_1}=\frac{\frac{B}{2}+X}{-x_2}=\frac{Y}{y_1}=\frac{Y}{y_2}=\frac{Z}{f}, \tag{1}$$

where f is the focal length of lens and the distance between the centers of the two lenses called the baseline B. The world coordinates(X, Y, Z)may be computed as follows:

$$X=\frac{B}{2}\left(\frac{x_1+x_2}{x_2-x_1}\right), \tag{2}$$

$$Y=\frac{B\cdot y_1}{x_1-x_2}=\frac{B\cdot y_2}{x_1-x_2}, \tag{3}$$

and

$$Z=\frac{B\cdot f}{x_1-x_2}. \tag{4}$$

Eqs. (2)-(4) indicate that if the baseline$B$ and focal length of lens f are known, and the corresponding image coordinates (x_1, y_1) and (x_2, y_2) can be determined, computing the coordinates (X, Y, Z) of P is a simple matter.

3. Modified Stereo Vision Model

In this section, we describe the principle of our proposed approach for improving the accuracy of depth estimation from two stereo images using a modified stereo vision model, especially suitable for variable focus cameras. First, we describe the

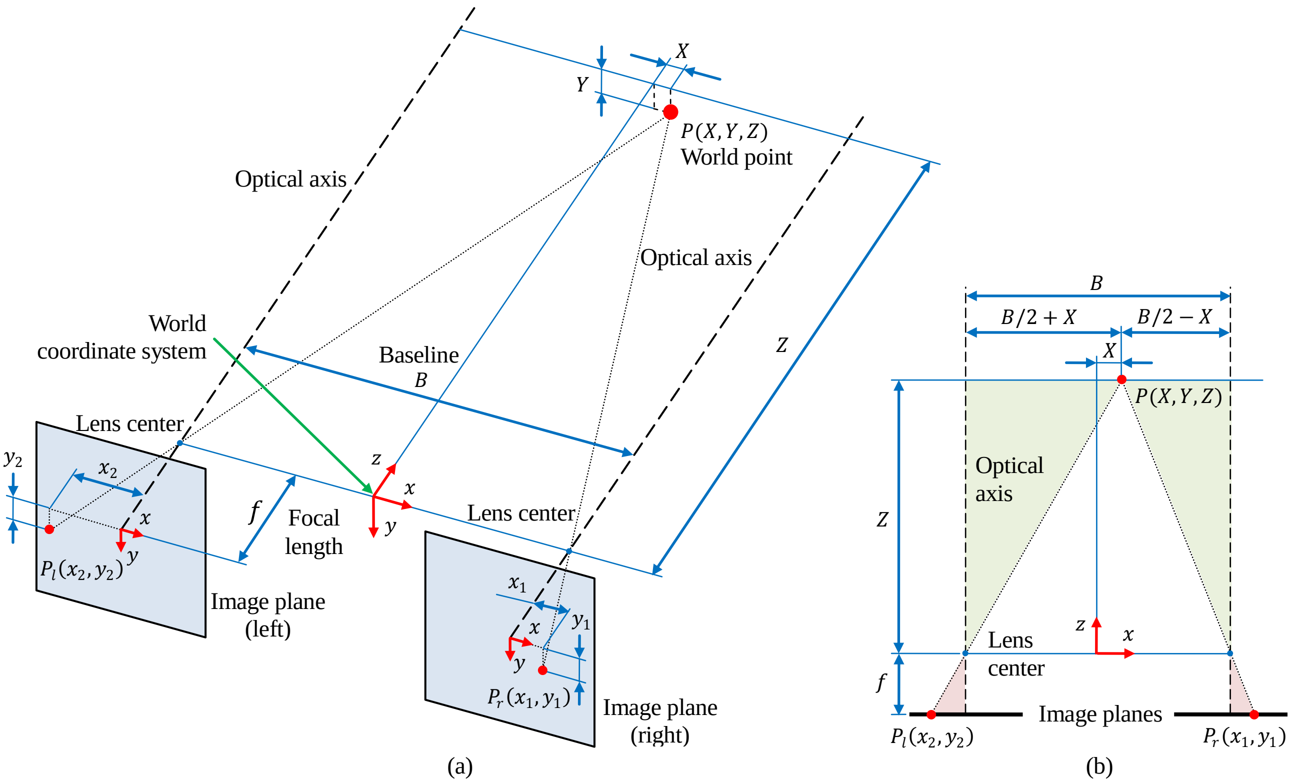
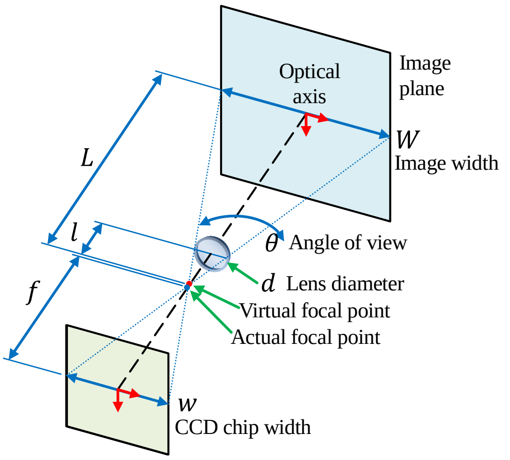


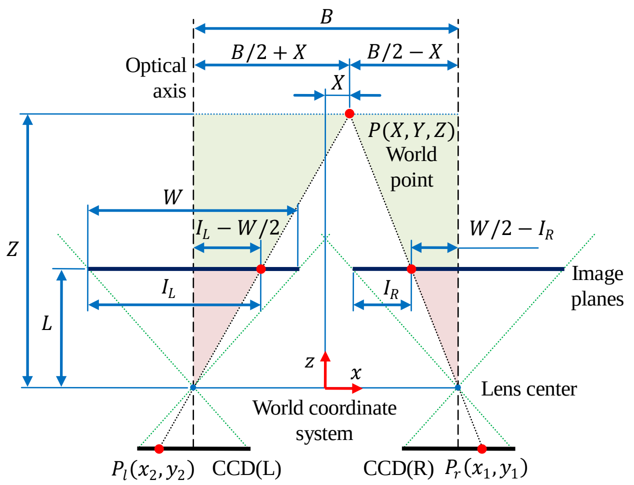
Fig. 1 Parallel axis geometry of conventional stereo vision: (a) 3-D view;(b) top view

modification of the geometry of conventional stereo vision. Then, we derive the equations for computing the coordinates (X, Y, Z) of P without the term focal length.

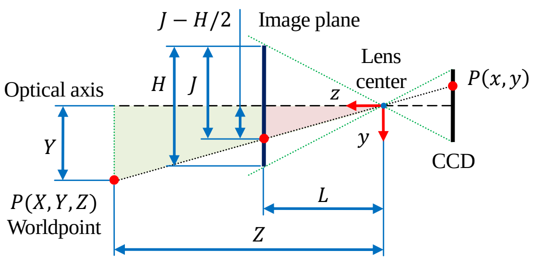
As stated above, an important problemexists in variable focus stereo visionis that the focal length of lens may notbe actually fixed. For this reason, using Eq. (4) will reduce the accuracy of depth dimension. In this paper, we propose a modified model of binocular stereo vision as shown in Fig. 2. Two cameras are separated by a distance$B$ in the x-direction and both optical axes are parallel. For convenience, the coordinate system centered between two cameras is called the world coordinate system.



(a)



(b)



(c)

Fig. 2 The modified model of binocular stereo vision: (a) schematic diagram of CCD imaging; (b) top view; (c) right view

As shown in Fig. 2(a), we can easily measure the distance$L$(in centimeters) between the image plane and the virtual focal point using a simple procedure as follows. First,by measuring the diameter$d$ and the angle of view θ of the lens, the distance l between lens and virtual focal point can be obtained. Then, if an arbitrary image plane is in front of the camera anditslocationis known, we can capture the image to measureits width. Consider the geometry shown in Fig. 2(a), the distance L between the image plane and virtual focal point can be determined.

From the similar triangles of imaging as shown in Figs. 2(b) and 2(c), we have

$$\frac{\frac{B}{2}-X}{\frac{W}{2}-I_R} = \frac{\frac{B}{2}+X}{I_L-\frac{W}{2}} = \frac{Y}{J-\frac{H}{2}} = \frac{Z}{L}, \tag{5}$$

where $I_L$and I_R are the horizontal positions and$J$ is the vertical position of the object in the two images (in pixels), W and $H$are the image width and height (in centimeters), respectively. In Eq. (5), we may note that only the units ofI_L,I_R, and J are in pixels. Here, we assumed that the image resolution is$M \times N$ (in pixels).After being convertedpixels to centimeters, Eq. (5) can be expressed in the form

$$\frac{\frac{B}{2}-X}{\frac{W}{2}-I_R\frac{W}{M}} = \frac{\frac{B}{2}+X}{I_L\frac{W}{M}-\frac{W}{2}} = \frac{Y}{J\frac{H}{N}-\frac{H}{2}} = \frac{Z}{L}. \tag{6}$$

Then, the world coordinates(X, Y, Z)may be computed as follows:

$$X = \frac{B(I_L + I_R - M)}{2(I_L - I_R)}, \tag{7}$$

$$Y = \frac{BMH(2J - N)}{2NW(I_L - I_R)}, \tag{8}$$

and

$$Z = \frac{BLM}{W(I_L - I_R)}. \tag{9}$$

Compared with Eq. (4), the term focal length$f$does not appear in Eq. (9).Finally, note that all coordinates obtained with Eqs. (7)-(9) are with respect to the world coordinate system centered between two cameras.

4. Experimental Results

In this paper, we designed the binocular vision system illustrated in Fig. 3, and the values of the parameters of the system, such as image resolution, image format, baseline length, etc., are shown in Table 1. Two identical CCD cameras are mounted in parallel on a platform with two rotatory degrees of freedom. We applied the proposed stereo algorithm to calculate the depth of the target centroid in two images illustrated in Fig. 4. The stereo images have a size of 640×480 pixels.A set of tabulated results are shown in Table 2. As a measure of accuracy of the depth computed using the proposed method in this paper, an error parameter defined as

$$\text{error} = \frac{\text{computed depth} - \text{actual depth}}{\text{actual depth}} \times 100\%. \quad (10)$$

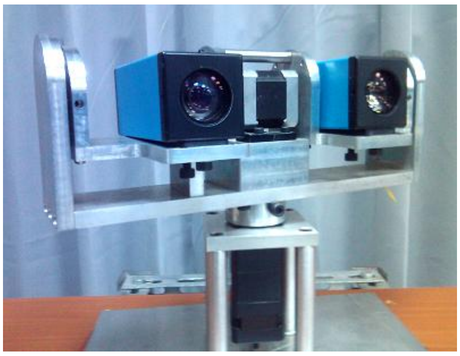


Fig. 3Photograph of the binocular vision system

Image resolution	640×480
Image format	RGB
Baseline length	20 cm
Angle of view	42.8434°
Distance between image plane and virtual focal point	0.4715 cm
Average computing time (s)	0.74 s

Table. 1Cameraparameters of the binocular vision system.

Actualdepth (cm)	Computed depth (cm) and error (%)	
	Conventional stereo	Modified stereo
80	75.5217(-5.60%)	79.0869(-1.14%)
90	84.5217(-6.09%)	88.5118(-1.65%)
100	93.7511(-6.25%)	98.1769(-1.82%)
110	103.2449(-6.14%)	108.1188(-1.71%)
120	110.9707(-7.52%)	116.2093(-3.16%)
130	119.9462(-7.73%)	125.6086(-3.38%)
140	129.4658(-7.52%)	135.5776(-3.16%)
150	138.2431(-7.84%)	144.7693(-3.49%)
160	146.9611(-8.15%)	153.8988(-3.81%)

Table. 2 The computed depth and error comparison of conventional and modified stereo models.

The experimental results show that the accuracy of computed depth by the new approach is better than the conventional method, namely, the new approach helps decrease the error of depth estimation about one half compared with the conventional method.

5. Conclusions

Most of the research works in stereo vision are focused on finding stereo correspondence methods and all of them assumed that the camera's focal length is known. Here, we have no doubt about the importance of the stereo correspondence problem in stereo vision.To date, however, variable focal length lens systems are becoming more and more popular, and therefore it is necessary to modify the conventional stereo vision model. In this paper, an efficient and accurate depth extraction approach based on a modified computational stereo model in binocular vision is presented. First, we use a simple procedure to measure the distance between the image plane and the virtual focal point. Then we can compute the coordinates of a world point without using the term focal length. The experimental results show that the proposed method in this paper has the ability to improve the accuracy of depth estimation.We can finally conclude that a slightly modified stereo model can extract the accurate depth information from stereo pairs. Besides, it is more important that this model is especially suitable for use in variable focal length systems.

ACKNOWLEDGEMENT

This work was supported by the National Science Council of Taiwan (R.O.C.) under grantnumber NSC 99-2221-E-0033-028 and the CYCU Distinctive Research Area project as grant CYCU-98-CR-ME.

REFERENCES

1. Li, Z. N. and Hu, G., "Analysis of Disparity Gradient Based Cooperative Stereo," IEEE Transactions on Image Processing, Vol. 5, No. 11, pp. 1943-1506, 1996.
2. Nitzan, D., "Three-Dimensional Vision Structure for Robot Applications," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No. 3, pp. 291-309, 1988.
3. Weng, J., Cohen, P., and Rebibo, N., "Motion and Structure Estimation from Stereo Image Sequences," IEEE Trans. on Robotics and Automation, Vol. 8, No. 3, pp. 362-382, 1992.
4. Marr, D. and Poggio, T. A., "Cooperative Computation of Stereo Disparity," Science, Vol. 194, No. 4262, pp. 283-287, 1976.
5. Barnard, S. T. and Fischler, M. A., "Computational Stereo," Computing Surveys, Vol. 14, No. 4, pp. 553-572, 1982.
6. Brown, M. Z., Burschka,D., and Hager,G. D.,"Advances in Computational Stereo," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, No. 8, pp. 993-1008, 2003.
7. Grosso, E. and Tistarelli, M., "Active/Dynamic Stereo Vision,"IEEE Trans. on Pattern Analysis and Machine Intelligence," Vol. 17, No. 9, pp. 868-879, 1995.
8. Yau, W. Y. and Wang, H., "Fast Relative Depth Computation for an Active Stereo Vision System," Real-Time Imaging, Vol. 5, pp. 189-202, 1999.
9. Ansari, M. E., Masmoudi, L., and Bensrhair, A., "A New Regions Matching for Color Stereo Images," Pattern Recognition Letters, Vol. 28, pp. 1679-1687, 2007.
10. Sengupta, S., "Effects of Unequal Focal Lengths in Stereo Imaging," Pattern Recognition Letters, Vol. 18, pp. 395-400, 1997.
11. Mizoguchi, N., Oku, H, and Ishikawa, M., "High-Speed Variable-Focus Optical System for Extended Depth of Field," IEEE International Symposium on Industrial Electronics, pp. 1668-1673, Seoul, Korea, July 5-8, 2009.
12. Ren, H. and Wu, S. T., "Variable-Focus Liquid Lens," Optics Express, Vol. 15, No. 10, pp. 5931-5936, 2007.
13. Gonzalea, R. C. and Woods, R. E., Digital Image Processing, Addison-Wesley Publishing Co. Inc., USA, 1992.

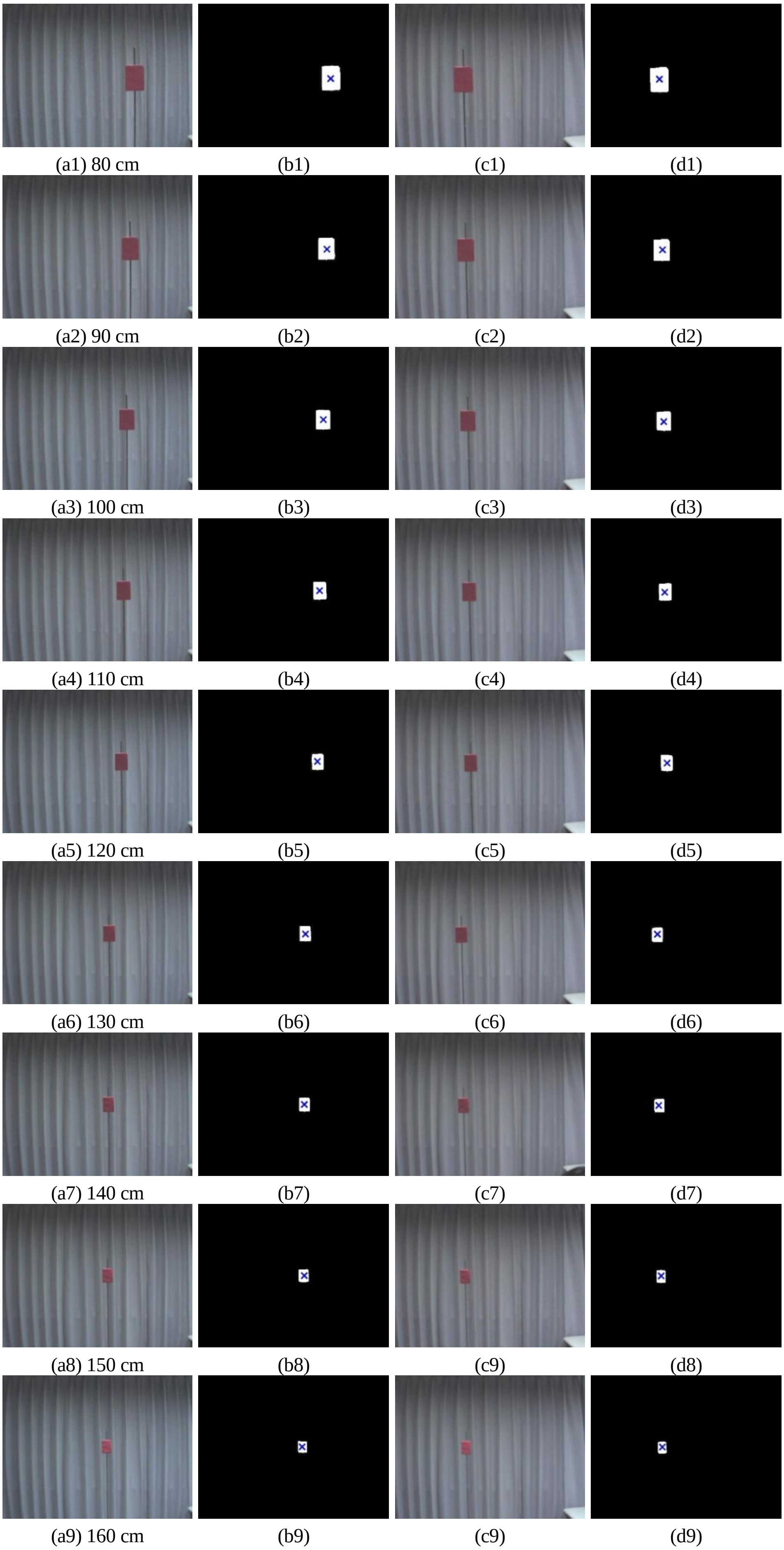


Fig. 4 The stereo image pairs at different distances: (a)captured byleft camera;(b) finding the centroid of the object in left image; (c) captured by right camera; (d) finding the centroid of the object in right image.