



# A new vision measurement method based on active object gazing

De Xu and Qingbin Wang

## Abstract

A new vision measurement system is developed with two cameras. One is fixed in pose to serve as a monitor camera. It finds and tracks objects in image space. The other is actively rotated to track the object in Cartesian space, working as an active object-gazing camera. The intrinsic parameters of the monitor camera are calibrated. The view angle corresponding to the object is calculated from the object's image coordinates and the camera's intrinsic parameters. The rotation angle of the object-gazing camera is measured with an encoder. The object's depth is computed with the rotation angle and the view angle. Then the object's three-dimensional position is obtained with its depth and normalized imaging coordinates. The error analysis is provided to assess the measurement accuracy. The experimental results verify the effectiveness of the proposed vision system and measurement method.

## Keywords

Stereo vision, active vision, calibration, depth estimation, 3-D measurement, visual system

Date received: 3 December 2016; accepted: 25 May 2017

Topic: Special Issue - Robot Perception for Manufacturing

Topic Editor: Stefano Ghidoni

Associate Editor: Emanuele Menegatti

## Introduction

Stereovision has been widely used in many applications.<sup>1–5</sup> For example, a stereo vision system fixed on welding robot is used by Chen et al.<sup>2</sup> to compute the position information of spatial seam, whose error is less than 0.3 mm. Xue et al.<sup>4</sup> reconstructed three-dimensional (3-D) multibubble trajectories and velocities effectively based on the captured images of multiple bubbles from two virtual views. A localization method of objects in 3-D scenes was developed using the cameras installed on collaborative robots, in which each robot has one camera.<sup>5</sup>

The traditional stereovision system consists of two cameras. A camera's model can be described with the pinhole model after its lens distortion is corrected or its lens distortion is negligible. The principle of stereovision, the well-known knowledge, is shortly introduced as follows. The measured scene point is on the line formed by the camera's optical center and the imaging point on the imaging plane. The 3-D position of the point is computed from the

intersection of two lines corresponding to the two cameras. The intrinsic and extrinsic parameters of the two cameras are calibrated in order to have the positions of optical centers and imaging points in the reference frame. The relative pose between the two cameras is stationary for the traditional stereovision system, which limits the common field of view and its flexibility in practice. It also results in the inconvenience of maintenance. Many researchers have been working to improve the performance of stereovision systems. For example, Tippetts et al.<sup>6</sup> provided an example

Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, China

### Corresponding author:

De Xu, Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing 100190, China.

Email: de.xu@ia.ac.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License

(<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of modifying an existing highly accurate stereo vision algorithm to increase its runtime performance while trying to limit the loss in accuracy. The field-programmable gate array (FPGA) was used in the implementation of local stereo methods to improve the real-time performance.<sup>7,8</sup> One of the solutions for the inconvenience problem in calibration is self-calibration. A measurement model for the stereo rig fixed beside the robot was developed based on the relative position of the end effector and target, which is linearly calibrated online with the robot's motions.<sup>9</sup>

The two images captured with a single camera at two viewpoints can also form stereovision.<sup>10</sup> It is the so-called motion-based stereovision. Its larger baseline is helpful to improve the measurement accuracy. But the relative pose of the camera from one to another viewpoint is not accurate, which has severe influence on the measurement accuracy.

It is known that the pinhole model of camera is from perspective projection. The values  $X/Z$  and  $Y/Z$  are equal to the coordinates on the imaging plane with the normalized focal length 1. The  $X$  and  $Y$  coordinates of object can be obtained if its depth  $Z$  is known. The depth is very important for 3-D measurement. Kinect cameras are widely used to provide depths images.<sup>11–14</sup> For example, a Kinect camera developed by Microsoft was used to capture color and depth images in order to recognize user gestures and human pose.<sup>11</sup> A real-time extraction method of surface patches was developed on Graphics Processing Unit (GPU) based on Kinect depth data in the study by Olesen et al.<sup>12</sup> However, the depth accuracy of Kinect cameras is low. Some researchers fuse stereo and Kinect data to obtain high resolution and high-quality depth estimation<sup>13</sup> or to provide an additional disparity map.<sup>14</sup> The color image combining depths is called RGBD (red, green, blue, depth) image, which can provide large convenience for object recognition or tracking.<sup>15–18</sup> For example, Chen et al. developed an FPGA-based RGBD imager, which is a trinocular stereo vision system to generate the composite color RGB and disparity data stream at video rate.<sup>15</sup> Teichman et al.<sup>16</sup> investigated the problem of segmenting and tracking deformable objects in RGBD data, which is available from commodity sensors such as the Asus Xtion Pro Live or Microsoft Kinect. Gedik and Alatan proposed an automated 3-D tracking algorithm based on fusion of vision and depth sensors via extended Kalman filter.<sup>17</sup> The RGBD videos of home-based exercise sessions for commonly applied shoulder and knee exercises were generated to recognize three main components of a physiotherapy exercise including the motion patterns, the stance knowledge, and the exercise object.<sup>18</sup> The small depth range and low accuracy of RGBD cameras limit its applications to the fields such as mobile robots navigation, object recognition, and object tracking.

Active vision systems can focus their field of view on the interested objects. New measurement model based on the stereovision principle and the rotation angles via active motions was developed by Xu et al.<sup>19</sup> The 3-D position of

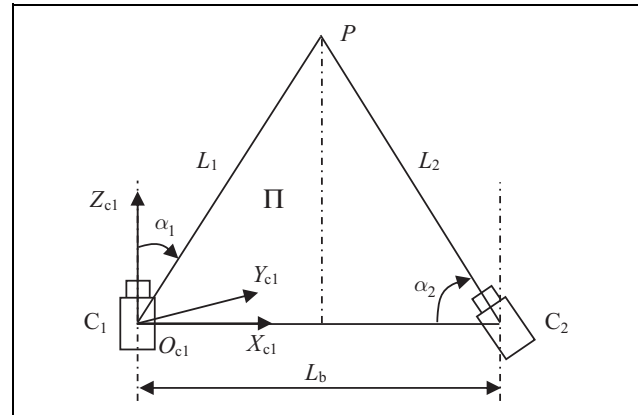


Figure 1. The principle of proposed vision system.

an object can be measured with two steps of active rotations of the two cameras to make the object at the central areas of images separately. The application convenience is improved since the intrinsic parameters of cameras are not required in measurement. The measurement area is enlarged by the active rotations. But the efficiency is very low because of two active rotations for each point measurement. The intrinsic parameters are easy to be calibrated in satisfactory accuracy, but the extrinsic parameters are difficult to be obtained in high accuracy. How to realize high accuracy measurement in a large range and ensure the convenience of application and maintenance is a valuable problem to be investigated.

The motivation of this work is to develop a new 3-D vision system with larger measurement range and good convenience of application and maintenance. The proposed vision system consists of two cameras. One is a passive camera, whose pose is fixed. It serves as a monitor camera to find and track objects in image space. The other is an active camera, whose orientation is actively changed. It works as an object-gazing camera to track the object in Cartesian space. The proposed vision system can be used in industrial manufacturing such as sheet metal fabrication to measure the size of large metal plate.

The rest of this article is organized as follows. The second section presents a new measurement method based on a fixed camera and an active camera. The calibration method of the proposed vision system is described in the third section. In the fourth section, the error analysis is given. Experiments and results are provided in the fifth section. The last section draws the conclusions.

## Vision system and measurement method

### Vision system

The principle of the proposed vision system is shown in Figure 1. The monitor camera  $C_1$  is fixed in position and orientation. Its intrinsic parameters are calibrated in advance. It is used to find and track objects in image space.

The object-gazing camera  $C_2$  is actively rotated around an axis parallel to  $Y_{c1}$ . Its intrinsic parameters are not needed since the tracked object is at the central area of image. Suppose the optical axes of two cameras are coplanar. The vision frame coincides with the frame of camera  $C_1$ . Its origin is the optical center of camera  $C_1$ .  $Z_{c1}$ -axis is selected to the optical axis direction toward scene.  $X_{c1}$ - and  $Y_{c1}$ -axis are the directions corresponding to the horizontal and vertical axes of image.

### Measurement method

In measurement, the camera  $C_2$  is actively rotated to make the horizontal image coordinate of point  $P$  at the central image area. The rotation angle  $\alpha_2$  is recorded with the encoder attached on the driven motor. On the other hand, the angle  $\alpha_1$  between the line  $C_1P$  and  $Z_{c1}$  is computed from the image coordinates of point  $P$  and the intrinsic parameters of camera  $C_1$ .

From the geometric relationship as shown in Figure 1, we have

$$z \tan\alpha_2 + z \tan\alpha_1 = L_b \quad (1)$$

where  $\alpha_1$  is the angle from  $Z_{c1}$  to  $C_1P$  projecting on the plane  $X_{c1}O_{c1}Z_{c1}$ ,  $\alpha_2$  is angle from  $-X_{c1}$  to  $C_2P$ .  $L_b$  is the distance from  $C_1$  to  $C_2$ , also known as baseline.  $z$  is the coordinate of point  $P$  in  $Z_{c1}$ -axis.

The depth  $z$  can be computed from (1) once the parameters  $L_b$ ,  $\alpha_1$ , and  $\alpha_2$  are known, as given in (2)

$$z = \frac{L_b}{\tan\alpha_1 + \tan\alpha_2} \quad (2)$$

The intrinsic parameters in pinhole model can be described as given in (3) if the lens distortion is corrected or negligible

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1c} \\ y_{1c} \\ 1 \end{bmatrix} \quad (3)$$

where  $(u_0, v_0)$  is the image coordinates of the principal point of camera  $C_1$ ,  $k_x$  and  $k_y$  are magnification factors of camera  $C_1$ .  $(x_{1c}, y_{1c})$  is the normalized imaging coordinates.

From (3), we have

$$\begin{cases} \tan\alpha_1 = x_{1c} = \frac{u - u_0}{k_x} \\ \tan\beta_1 = y_{1c} = \frac{v - v_0}{k_y} \end{cases} \quad (4)$$

where  $\beta_1$  is the angle from  $Z_{c1}$  to  $C_1P$  projecting on the plane  $Y_{c1}O_{c1}Z_{c1}$ .

The depth  $z$  is computed with (2) after  $\alpha_1$  and  $\alpha_2$  are obtained. Then, the coordinates  $x$  and  $y$  are computed as given in (5). Up to now, the 3-D position  $(x, y, z)$  of point  $P$  is obtained

$$\begin{cases} x = x_{1c}z \\ y = y_{1c}z \end{cases} \quad (5)$$

It can be found that the object-gazing camera  $C_2$  is very important to compute the object's depth. It need be rotated to aim the measured object. Therefore, only one object's depth can be measured for one time object gazing. Combining with the object's depth, its other coordinates in the vision frame are computed from its image coordinates and the intrinsic parameters of monitor camera  $C_1$ .

### Calibration

Considering the error of the initial zero position of the direction of camera  $C_2$ , the fixed error  $\delta_2$  of  $\alpha_2$  is introduced into the calibration. Thus, the calibration is to estimate the parameters  $L_b$  and  $\delta_2$ .

When the fixed error  $\delta_2$  of  $\alpha_2$  is introduced, formula (1) is rewritten as

$$z \tan(\alpha_2 + \delta_2) + z \tan\alpha_1 = L_b \quad (6)$$

The value range of  $\alpha_2$  is  $(0, \Pi)$ . In the case  $\alpha_2 \neq \Pi/2$ , (6) is expanded and arranged to (7). In the case  $\alpha_2 = \Pi/2$ , (6) is changed to (8)

$$L_b \tan\alpha_2 + L_b \tan\delta_2 + z(\tan\alpha_2 - \tan\alpha_1) \tan\delta_2 = z + z \tan\alpha_1 \tan\alpha_2 \quad (7)$$

$$L_b + z \tan\delta_2 = z \tan\alpha_1 \quad (8)$$

Formulae (7) and (8) are employed to calibrate the parameters  $L_b$  and  $\delta_2$  using multiple given calibration points with known  $z$  values.  $\alpha_1$ ,  $\alpha_2$ , and  $z$  are known variables in (7). Formula (7) is a linear equation with three unknown variables  $L_b$ ,  $L_b \tan\delta_2$ , and  $\tan\delta_2$  if  $L_b \tan\delta_2$  is taken as one new variable. In (8),  $\alpha_2 = \Pi/2$ ,  $\alpha_1$  and  $z$  are known variables. Formula (8) is a linear equation with two unknown variables  $L_b$  and  $\tan\delta_2$ . Thus, three calibration points at least can form three equations as given in (7) or (8) at least.  $L_b$ ,  $L_b \tan\delta_2$ , and  $\tan\delta_2$  can be solved using linear least square method. Of course, more calibration points are helpful to improve the calibration accuracy.

### Error analysis

From (4), we have  $\tan\alpha_1 = x_{1c}$ . In the following error analysis,  $\tan\alpha_1$  is taken as one original variable and replaced by  $x_{1c}$ . From the difference of (2), we have

$$dz = \frac{dL_b}{x_{1c} + \tan\alpha_2} - \frac{L_b [dx_{1c} - (1/\sin^2\alpha_2) d\alpha_2]}{(x_{1c} + \tan\alpha_2)^2} \quad (9)$$

Submitting (2) into (9) to simplify, then

$$dz = \frac{z}{L_b} dL_b - \frac{z^2}{L_b} dx_{1c} + \frac{z^2}{L_b \sin^2\alpha_2} d\alpha_2 \quad (10)$$

In addition, (11) is derived from (5)

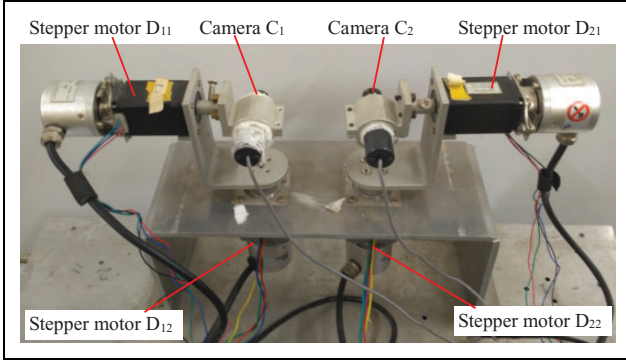


Figure 2. Experiment system.

$$\begin{cases} dx = z dx_{1c} + x_{1c} dz \\ dy = z dy_{1c} + y_{1c} dz \end{cases} \quad (11)$$

The relative errors to  $z$  are derived from (10) and (11)

$$\frac{dz}{z} = \frac{1}{L_b} dL_b - \frac{z}{L_b} dx_{1c} + \frac{z}{L_b \sin^2 \alpha_2} d\alpha_2 \quad (12)$$

$$\begin{cases} \frac{dx}{z} = dx_{1c} + x_{1c} \frac{dz}{z} \\ \frac{dy}{z} = dy_{1c} + y_{1c} \frac{dz}{z} \end{cases} \quad (13)$$

Generally,  $k_x$  and  $k_y$  are about 1000.  $dx_{1c} = 1/1000$  if the error in image coordinates is 1 pixel.  $d\alpha_2$  is the resolution of stepper motor, and it is  $0.9^\circ$  before subdivision. It is about  $0.05625^\circ$  after 16 subdivision, that is,  $9.82 \times 10^{-4}$  rad. Considering the case,  $L_b = 200$  mm,  $dL_b = 4$  mm, and  $z = 1000$  mm. In this case,  $\alpha_2$  is about  $80^\circ$  or 1.3963 rad. Submitting these data into (10), we have the error  $dz = 30.6$  mm and the relative error  $dz/z = 3.06\%$ . The term  $z dL_b/L_b = 20$  mm contributes the main part in the error  $dz$ . Assume  $dy_{1c} = 1/1000$ ,  $u - u_0 = 300$  pixel, and  $v - v_0 = 200$  pixel, then  $x_{1c} = 0.3$  and  $y_{1c} = 0.2$ . Submitting then into (11), we have the errors  $dx = 10.2$  mm and  $dy = 7.1$  mm. Then the relative errors  $dx/z = 0.10\%$  and  $dy/z = 0.07\%$ . In this case,  $z = 2000$  mm,  $dL_b = 6$  mm, and the other conditions are as the same as ones above. In this case,  $\alpha_2$  is about  $85^\circ$  or 1.4835 rad. In this case, the computed errors and relative errors are as follows.  $dz = 99.8$  mm,  $dx = 31.9$  mm,  $dy = 22.0$  mm,  $dz/z = 4.99\%$ ,  $dx/z = 1.60\%$ , and  $dy/z = 1.10\%$ .

## Experiments and results

### Experiment system

An experiment system was developed with two cameras driven by stepper motors, as shown in Figure 2. Each camera was driven by two stepper motors. One was used to yaw the camera and the other was used to pitch it. In

experiments, the stepper motors  $D_{11}$  and  $D_{21}$  were fixed to keep the two cameras  $C_1$  and  $C_2$  on a horizontal plane. In the experiments with the proposed method, the camera  $C_1$  was stationary and the camera  $C_2$  was yawed by the stepper motor  $D_{22}$  to gaze objects. In the comparing experiments with the traditional stereovision, the two cameras  $C_1$  and  $C_2$  were fixed to the given pose. The camera's type was OBK-2010CHEW with  $1/3''$  CCD. Its image size was  $720 \times 576$  pixels, and it was resized to  $400 \times 300$  pixels.

### Calibration results

The cameras were calibrated with OpenCV for the traditional stereovision. The cameras were described with the pinhole model and Brown lens model. The results were as follows.  $M_{in1}$  and  $M_{in2}$  were the intrinsic parameter matrixes, and  $k_1$  and  $k_2$  were the lens distortion factors of the two cameras  $C_1$  and  $C_2$ , respectively.  ${}^1T_2$  was the extrinsic parameter matrix, that is, the relative pose of the camera  $C_2$  to the camera  $C_1$

$$M_{in1} = \begin{bmatrix} 421.71 & 0 & 201.27 \\ 0 & 424.68 & 155.17 \\ 0 & 0 & 1 \end{bmatrix} \quad (14)$$

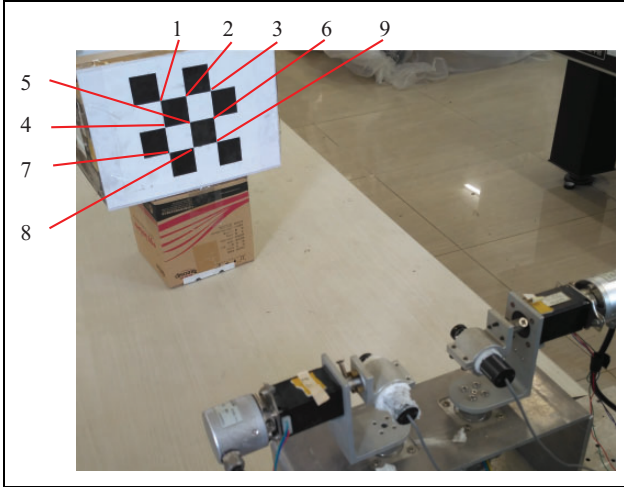
$$k_1 = [-0.3691 \quad 0.2223 \quad -0.0002 \quad 0.0001 \quad -0.1047] \quad (15)$$

$$M_{in2} = \begin{bmatrix} 422.67 & 0 & 196.20 \\ 0 & 425.58 & 145.12 \\ 0 & 0 & 1 \end{bmatrix} \quad (16)$$

$$k_2 = [-0.3730 \quad 0.2813 \quad 0.0008 \quad 0.0001 \quad -0.2299] \quad (17)$$

$${}^1T_2 = \begin{bmatrix} 0.9999 & -0.0091 & 0.0126 & 77.1391 \\ 0.0090 & 0.9999 & 0.0053 & -7.9497 \\ -0.0127 & -0.0052 & 0.9999 & 2.3782 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (18)$$

A chessboard pattern was put at different positions in front of the vision system, as shown in Figure 3. The camera  $C_1$  was stationary and the camera  $C_2$  was yawed by the stepper motor  $D_{22}$ . Firstly, the vision system worked as the traditional stereovision after the camera  $C_2$  was yawed to the given pose. The 3-D position of the center of chessboard pattern was measured with the traditional stereovision. Then, the camera  $C_2$  was yawed to gaze the center of chessboard pattern. Its yawing angle was recorded. The collected data in the baseline calibration is listed in Table 1. The parameters  $L_b$  and  $\tan \delta_2$  were computed with the method in the third section;  $L_b = 77.09$  mm,  $\tan \delta_2 = 0.0133$ , and  $\delta_2 = 0.762^\circ$ .



**Figure 3.** Calibration scene.

**Table 1.** Data in baseline calibration.

No.	$z$ (mm)	$\tan\alpha_1 (x_{1c})$	$\text{ctan}\alpha_2$
1	485.9	0.1018	0.0673
2	486.5	0.2635	-0.0876
3	796.4	0.2353	-0.1219
4	805.9	0.0229	0.0838
5	1008.5	0.0162	0.0724
6	998.5	0.1624	-0.0705

### Measurement experiment

In experiments, the nine corners on the chessboard pattern as shown in Figure 3 were measured. In the measurement with the proposed method, the pattern was taken as an object. Its central corner's depth was considered to be the pattern's depth. It was also used for all nine corners. It was computed from the baseline and the angles when the camera  $C_2$  was yawed to gaze the central corner. In the measurement with the traditional stereovision method, all nine corners were separately computed with the intrinsic and extrinsic parameters of the cameras and the image coordinates. The measured 3-D positions of the corners were listed in Table 2. The actual depth was manually measured with a ruler.

In the measurement with the traditional stereovision method, the corners numbered 1, 4, and 7 were out of the field of view of the camera  $C_2$ . Their positions were absence, which are denoted with “/” as shown in Table 2. In the measurement with the proposed method, all corners were in the two cameras' fields of view and their positions were obtained.

It can be found from Table 2 that the relative depth errors were 0.3%, 1.2%, 3.6%, and 7.5% in the four groups of experiments with the proposed method. It is coincidence with the error analysis in the fourth section. In the proposed method, all depth errors were positive. The possible reason was that the rotation axis was at the left side of the optical axis, which resulted in the positive error  $dL_b$ . In the first

group of experiments with the traditional stereovision method, the relative depth errors were from -9.5% to -7.9% for the corners in the field of view. They were from 0.2% to 0.7% for the second group of experiments with the traditional stereovision method. They were from -3.2% to -1.6% for the third group of experiments with the traditional stereovision method. They were from -1.7% to -0.3% for the fourth group of experiments with the traditional stereovision method. The stereovision method could not measure three points in the first group experiment since they were out of the view field of the camera  $C_2$ .

On the one hand, as shown in (12) in the error analysis section, the measurement relative error will increase with the depth increment. On the other hand, it is difficult to ensure the camera's rotation axis passes through the camera's projection center or optical axis. It is sure that the offset of the camera's rotation axis from its optical axis is a resource to result in the error  $dL_b$ . It can be found from Table 1 that the proposed vision system was calibrated when the target was at the positions with depths from 485 mm to 1000 mm. The calibrated parameters of the proposed vision were much more suitable to this depth range. The target's depths in the first and second group of experiments were 350 mm and 1000 mm, which were near to the target's depths in calibration. Therefore, the measurement results in the first and second group of experiments were more accurate than others.

A pair of captured images in the measurement with the proposed method is given in Figure 4. The measured positions of the corners are shown in Figure 5. It can be found from Figure 5 and Table 2 that the measurement accuracy with the proposed method was as same level as one with the stereovision method. But the proposed method can work in a larger range.

### Measurement experiment for large size component

In industrial manufacturing, the components to be fabricated sometimes are of large size. For example, the metal plate in sheet metal fabrication may have large size up to 2 m long and 1 m wide. The traditional stereovision cannot have such large a common field of view to measure the metal plate's size when it is near to the vision system.

As the imitation of the components in sheet metal fabrication, three plate components in office were used to test the measurement with the developed vision system. Figure 6 shows the images in the process to measure the first long plate. Figure 6(a) is the image captured by the camera  $C_1$ . Figure 6(b), 6(c), and 6(d) are the images captured by the camera  $C_2$  at the initial pose, the pose to measure far point, and the pose to measure near point, respectively. It can be found from Figure 6 that the camera  $C_2$  cannot capture the whole image of the long plate at above poses. In other words, it is difficult to measure the size of the long plate with the traditional stereovision system.

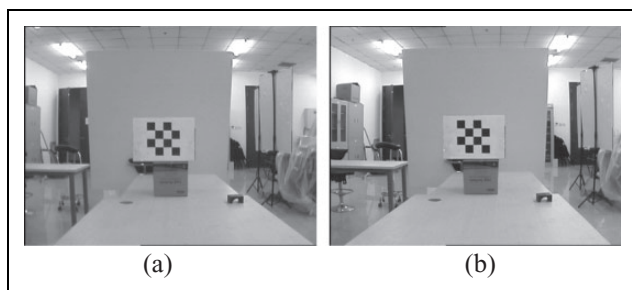
In the measurement procedure, the camera  $C_2$  was firstly yawed to gaze the far point of the long plate, and the



**Table 2.** Measured 3-D positions of the pattern corners.<sup>a</sup>

No	$x_{1c}$	$y_{1c}$	$\text{ctan}\alpha_2$	Stereovision (mm)			Proposed method (mm)			Actual z (mm)
				X	y	z	x	y	z	
1-1	-0.2577	-0.1511	0.4860	/	/	/	-90.4	-53.0	350.9	350
1-2	-0.1363	-0.1491	0.3714	-44.0	-47.9	322.4	-47.8	-52.3		
1-3	-0.0148	-0.1454	0.2554	-4.8	-46.7	321.3	-5.2	-51.0		
1-4	-0.2597	-0.0300	0.4883	/	/	/	-91.1	-10.5		
1-5	-0.1366	-0.0272	0.3714	-43.8	-8.7	320.5	-47.9	-9.6		
1-6	-0.0158	-0.0252	0.2554	-5.0	-8.0	320.5	-5.5	-8.8		
1-7	-0.2605	0.0918	0.4868	/	/	/	-91.4	32.2		
1-8	-0.1379	0.0930	0.3714	-43.6	29.5	316.6	-48.4	32.6		
1-9	-0.0165	0.0950	0.2554	-5.2	30.4	319.4	-5.8	33.3		
2-1	-0.0402	-0.0466	0.1276	-41.2	-47.7	1027.0	-41.5	-48.1	1032.3	1020
2-2	-0.0005	-0.0461	0.0888	-0.9	-47.2	1024.7	-0.5	-47.6		
2-3	0.0392	-0.0454	0.0497	40.1	-46.3	1023.7	40.4	-46.9		
2-4	-0.0407	-0.0064	0.1282	-41.5	-6.5	1022.2	-42.0	-6.6		
2-5	-0.0007	-0.0060	0.0888	-0.7	-6.1	1022.7	-0.8	-6.2		
2-6	0.0389	-0.0057	0.0497	39.8	-5.8	1022.0	40.1	-5.9		
2-7	-0.0408	0.0328	0.1270	-41.5	33.9	1024.3	-42.1	33.9		
2-8	-0.0012	0.0339	0.0869	-1.2	34.8	1023.3	-1.2	35.0		
2-9	0.0388	0.0347	0.0484	39.7	35.3	1021.6	40.1	35.8		
3-1	-0.0462	-0.0260	0.1060	-73.7	-41.4	1594.3	-77.5	-43.7	1678.6	1620
3-2	-0.0197	-0.0257	0.0800	-31.4	-40.8	1589.5	-33.0	-43.1		
3-3	0.0070	-0.0254	0.0541	11.2	-39.8	1570.2	11.8	-42.7		
3-4	-0.0465	0.0002	0.1060	-73.6	0.5	1583.3	-78.0	0.4		
3-5	-0.0200	0.0004	0.0793	-31.9	0.7	1588.4	-33.6	0.7		
3-6	0.0069	0.0007	0.0541	10.7	1.2	1571.6	11.6	1.1		
3-7	-0.0464	0.0263	0.1079	-73.7	41.9	1586.3	-77.8	44.2		
3-8	-0.0201	0.0271	0.0812	-31.9	43.0	1583.0	-33.7	45.4		
3-9	0.0068	0.0274	0.0547	10.7	43.2	1568.2	11.4	46.0		
4-1	-0.0354	-0.0166	0.0838	-74.1	-34.6	2102.3	-80.8	-38.0	2281.9	2120
4-2	-0.0157	-0.0162	0.0629	-33.2	-34.4	2114.4	-35.9	-37.0		
4-3	0.0044	-0.0161	0.0434	9.7	-33.5	2084.3	10.1	-36.8		
4-4	-0.0354	0.0034	0.0831	-74.7	7.4	2112.5	-80.7	7.9		
4-5	-0.0158	0.0036	0.0629	-33.1	7.5	2112.8	-36.0	8.2		
4-6	0.0044	0.0037	0.0421	9.4	8.0	2095.7	10.1	8.5		
4-7	-0.0354	0.0227	0.0838	-73.8	47.5	2090.1	-80.7	51.7		
4-8	-0.0158	0.0233	0.0629	-32.9	48.9	2093.3	-36.0	53.2		
4-9	0.0044	0.0234	0.0440	9.2	48.9	2083.1	10.0	53.3		

<sup>a</sup>"/" means the value cannot be obtained.

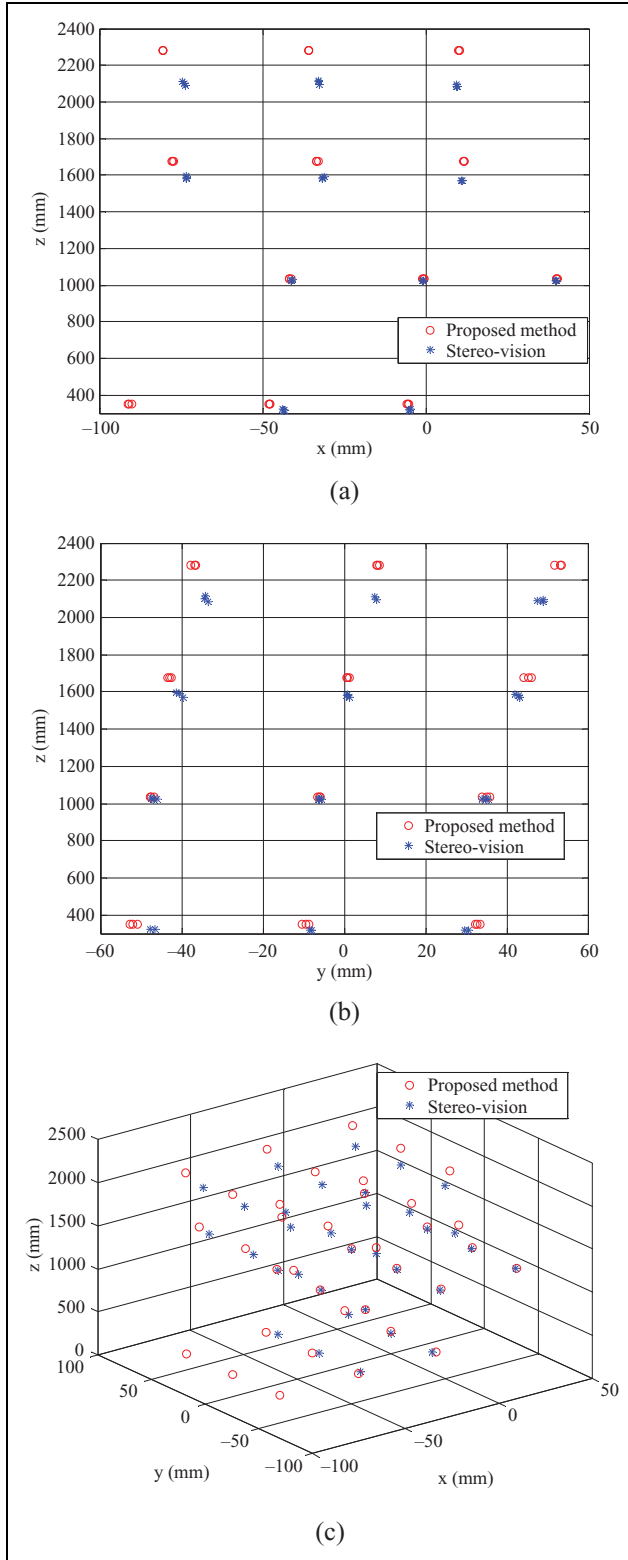
**Figure 4.** Captured images: (a) camera  $C_1$  and (b) camera  $C_2$ .

coordinates of the far point were computed with the proposed method. Secondly, the camera  $C_2$  was yawed to gaze the near point of the long plate, and the coordinates of the near point were computed. Then the length of the long plate was calculated from the far and near points. In the

experiments, three components were measured. The other two components were shown in Figure 7. Their measured lengths were 1207.2 mm, 529.6 mm, and 399.3 mm, respectively. Their actual lengths were 1400 mm, 500 mm, and 400 mm, respectively. The actual lengths were manually measured with a ruler. The measurement error was large for the first long plate component since its smooth corners reduced the match accuracy of the corresponding feature points. The measurement errors were much small for the second and third plate components since their sharp corners were helpful to ensure the match accuracy of the corresponding feature points.

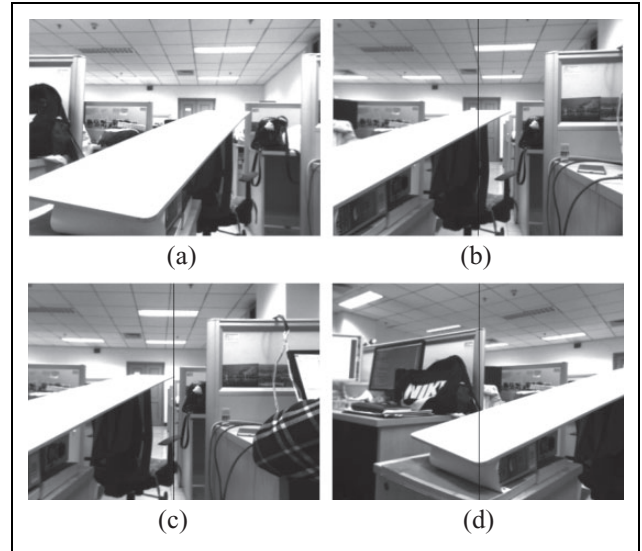
## Conclusion

The main contribution of this work is the developed new active vision system. It monitors objects with a fixed

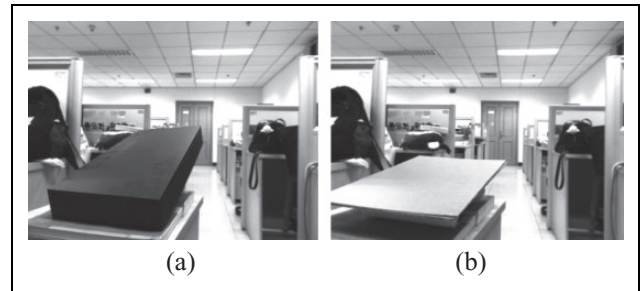


**Figure 5.** Measured positions of the corners: (a) XOZ, (b) YOZ, and (c) 3-D.

camera and gazes them with an active yawing camera. The object’s depth is computed from the baseline, image coordinates in the fixed camera, and the yawing angle of the



**Figure 6.** Captured images of first long plate component: (a) camera  $C_1$ , (b) camera  $C_2$  at initial pose, (c) camera  $C_2$  at the pose to measure far point, and (d) camera  $C_2$  at the pose to measure near point.



**Figure 7.** Captured images of other two plate components: (a) image of second component captured by camera  $C_1$  and (b) image of third component captured by camera  $C_1$ .

object-gazing camera. Then the object’s other coordinates in Cartesian space are calculated with the depth and the normalized imaging coordinates. Experimental results verify the effectiveness of the proposed active vision system and measurement method.

The proposed vision system has much larger measurement range, more flexibility, and better convenience of application and maintenance than the traditional stereovision system.

In future, we will focus on the improvement of measurement accuracy with the proposed vision system. More factors such as the resolution of the cameras, the lens distortion, the position of rotation axis of camera  $C_2$ , and the chessboard pattern<sup>20</sup> will be taken into account.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China under grants 61421004, 61227804, and 61403378 and the National High Technology Research and Development Program of China (863) under grant 2015AA042307.

## References

1. Tippetts BJ, Lee DJ, Archibald JK, et al. Dense disparity real-time stereo vision algorithm for resource-limited systems. *IEEE Trans Circuits Syst Video Technol* 2011; 21(10): 1547–1555.
2. Chen XZ, Huang YM, and Chen SB. Model analysis and experimental technique on computing accuracy of seam spatial position information based on stereo vision for welding robot. *Ind Robot Int J* 2012; 39(4): 349–356.
3. Chiang MH, Lin HT, and Hou CL. Development of a stereo vision measurement system for a 3D three-axial pneumatic parallel mechanism robot arm. *Sensors* 2011; 11(2): 2257–2281.
4. Xue T, Qu L, and Wu B. Matching and 3-D reconstruction of multibubbles based on virtual stereo vision. *IEEE Trans Instrum Meas* 2014; 63(6): 1639–1647.
5. Lins RG, Givigi SN, and Gardel Kurka PR. Vision-based measurement for localization of objects in 3-D for robotic applications. *IEEE Trans Instrum Meas* 2015; 64(11): 2950–2958.
6. Tippetts B, Lee DJ, Lillywhite K, et al. Efficient stereo vision algorithms for resource-limited systems. *JRTIP* 2015; 10(1): 163–174.
7. Fife WS and Archibald JK. Improved census transforms for resource-optimized stereo vision. *IEEE Trans Circuits Syst Video Technol* 2013; 23(1): 60–73.
8. Jin S, Cho J, Pham XD, et al. FPGA Design and implementation of a real-time stereo vision system. *IEEE Trans Circuits Syst Video Technol* 2010; 20(1): 15–26.
9. Shen Y, Xu D, Tan M, et al. Mixed visual control method for robots with self-calibrated stereo rig. *IEEE Trans Instrum Meas* 2010; 59(2): 470–479.
10. Olson CF and Abi-Rached H. Wide-baseline stereo vision for terrain mapping. *Mach Vis Appl* 2010; 21(5): 713–725.
11. Sharma K, Moon I, and Kim SG. Depth estimation of features in video frames with improved feature matching technique using Kinect sensor. *Opt Eng* 2012; 51(10): 1–11.
12. Olesen SM, Lyder S, Kraft D, et al. Real-time extraction of surface patches with associated uncertainties by means of Kinect cameras. *JRTIP* 2015; 10(1): 105–118.
13. Zhang S, Wang C, and Chan SC. A new high resolution depth map estimation system using stereo vision and Kinect depth sensing. *J Signal Proc Syst Signal Image Video Technol* 2015; 79(1): 19–31.
14. Williem, Tai YW and Park IK. Accurate and real-time depth video acquisition using Kinect-stereo camera fusion. *Opt Eng* 2014; 53(4): 1–9.
15. Chen L, Jia Y, and Li M. An FPGA-based RGBD imager. *Mach Vis Appl* 2012; 23(3): 513–525.
16. Teichman A, Lussier JT, and Thrun S. Learning to segment and track in RGBD. *IEEE Trans Autom Sci Eng* 2013; 10(4): 841–852.
17. Gedik OS and Alatan AA. 3-D rigid body tracking using vision and depth sensors. *IEEE Trans Cyber* 2013; 43(5): 1395–1405.
18. Ar I and Akgul YS. A computerized recognition system for the home-based physiotherapy exercises using an RGBD camera. *IEEE Trans Neural Syst Rehabil Eng* 2014; 22(6): 1160–1171.
19. Xu D, Li YF, Tan M, et al. A new active visual system for humanoid robots. *IEEE Trans Syst Man Cyber B Cyber* 2008; 38(2): 320–330.
20. Furferi R, Governì L, Volpe Y, et al. Design and assessment of a machine vision system for automatic vehicle wheel alignment. *Int J Adv Robot Syst* 2013; 10: 242.