IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Target distance measurement method using monocular vision

ISSN 1751-9659
Received on 7th October 2019
Revised 4th April 2020
Accepted on 18th June 2020
E-First on 2nd October 2020
doi: 10.1049/iet-ipr.2019.1293
www.ietdl.org

Mao Jiafa[1] ✉, *Huang Wei*[1], *Sheng Weiguo*[2]

[1]*College of Computer Science and Technology, Zhejiang University of Technology, Hang Zhou, Zhejiang 310023, People's Republic of China*
[2]*Department. of Computer Science, Hangzhou Normal University, Hangzhou, Zhejiang 311121, People's Republic of China*
✉ *E-mail: maojiafa@zjut.edu.cn*

Abstract: Most existing machine vision-based location methods mainly focus on the spatial positioning schemes using one or two cameras along with non-vision sensors. To achieve an accurate location, both schemes require processing a large amount of data. In this study, the authors propose a novel method, which requires much less amount of data to be processed for measuring target distance using monocular vision. Based on the geometric model of camera imaging, the parameters of the camera (such as camera's focal length and equivalent focal length.), as well as the principle of analogue signal being transformed into a digital signal, the authors derive the relationship among the target distance, field of view, equivalent focal length and camera resolution. Experimental results show that the proposed method can effectively and accurately achieve the target distance measurement.

1 Introduction

In the past two decades, robots have been widely used in various fields including monitoring, industrial automation production, visual navigation, automatic image interpretation, human–computer interaction and virtual reality [1]. Machine vision is to simulate the visual function of human eyes using computers, whose objective is to extract information from images or image sequences and then perform morphological and motion recognition for three-dimensional (3D) scenes and objects in the real world [2–4].

Employing robots to the fields, such as playing football, cleaning room, looking after children and patients, is now highly expected [5–8]. The performance of robots depends largely on the technology of machine vision. Human being utilises vision to obtain external information, while the machine vision system relies on cameras to collect images. Such images are used to control the behaviour of robots. Due to the high flexibility and adaptability, machine vision technology has been widely employed. In machine vision, the acquisition of target distance is one of the most difficult problems in the fields such as autonomous navigation, 3D scene reconstruction, vision-based measuring and industrial automation [9].

1.1 Prior work

The 3D location techniques are vital in machine vision and have been widely studied [10–12]. Doh *et al.* [13] introduced a simultaneous localisation and mapping (SLAM) algorithm for topological maps with dynamics. Yang *et al.* [14] proposed dynamic RGB-D encoder SLAM for a differential-drive robot. Fernandez *et al.* [15] devised the appearance-based approach to hybrid metric-topological SLAM. These state-of-the-art SLAM methods [13–15] could have high accuracy localisation capabilities and impressive mapping effects. However, they assume that the operating environment is static, thus limiting their application in the real dynamic world. Saxena *et al.* [16] exploited the Markov random field to learn 3D scene structure using a single still image. The learnt model captures information form depth cues as well as the relationships between image parts. Haim *et al.* [17] proposed to employ a phase-coded aperture camera for depth estimation. They use a fully convolutional neural network to estimate depth maps. Liu *et al.* [18] employed the fully convolutional network (FCN) architecture for monocular depth estimation. Eigen *et al.* [19] introduced a deep neural network for depth estimation that relies on depth cues in the RGB image. He *et al*. [20] devised a method to generate synthetic varying focal-length data set from fixed-focal-length data sets. They also implemented a simple yet effective method to fill the holes in the generated images. Zhang *et al.* [21] developed a deep hierarchical guidance and regularisation (HGR) learning framework for end-to-end monocular depth estimation.

The stereo vision system of our human being is derived from the difference of viewing angle between the left and right eyes. This is important source of human and other animals perceiving depth and structured information [22, 23]. Yang *et al.* [24] proposed a stereo matching method for binocular camera in gesture recognition. After estimating the parallax through the relative information between two cameras, they obtained the distance of the target.

The camera calibration technology has been widely used in monocular distance perception [25–32]. As one of the fundamental problems in machine vision, camera calibration is to determine internal and external parameters of cameras by using image features and the corresponding 3D features. Camera calibration has been extensively studied in machine vision, and several classical calibration methods have been designed. Abdel-Aziz and Karara [33] first proposed the direct linear transformation method. Tsai [34] proposed a two-level calibration method called the radial uniform constraint. Zhang [35] designed a relatively simple and flexible calibration method. The key component of traditional calibration methods is to establish the physical coordinates of the target in a static environment. As a result, these methods are difficult to be applied to the mobile robot vision. To address this issue, the camera calibration method based on active vision has also been proposed in [36]. This method estimates depth information using multiple images. This method, however, increases the amount of data and reduces the calculation speed, which is crucial for real-time positioning in machine vision.

The self-calibration method relies only on the relationship between the corresponding points of multiple images. Pollefeys *et al.* [37] devised a practical camera self-calibration method with variable internal parameters. Ferran [38] proposed a linear self-calibration method for planar motion. This method uses multiple images to achieve distance measurement could also increase the data amount to be processed and reduce the speed of 3D

© The Institution of Engineering and Technology 2020

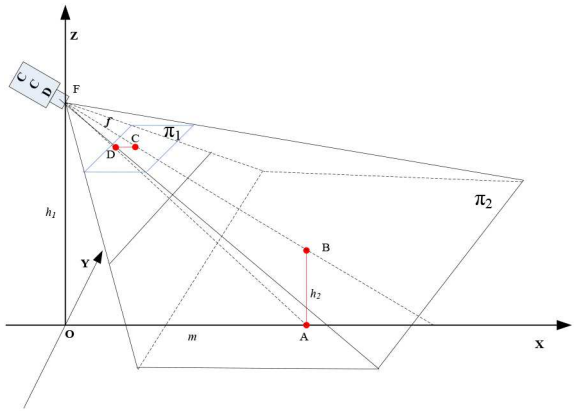

Fig. 1 *Illustration of camera imaging when the top of the target is in the midpoint of vision*

positioning, which restricts its application in real-time positioning of machine vision.

In [39–41], the authors designed the fish-group target information acquisition platform to achieve multi-target tracking and positioning by using a single-camera-plane depth acquisition. The image in the plane mirror in these methods is used to measure the target depth information. These methods can only be used in specific environments. Laurel *et al.* [42] used a single camera with double spotlights to capture the fish and their shadow information. This method is able to accurately locate the 3D fish position in most sentiment. However, it is not suitable for target depth measurement in mobile environments.

1.2 Motivations

The methods of depth information extraction for external targets can be roughly categorised into two categories. The first one is based on non-visual sensors, while the second one is based on computer vision. Here, we mainly focus on the vision-based positioning scheme, which mainly includes: (i) target distance estimation method from a single image based on deep learning, (ii) the binocular perception depth method, (iii) the method of camera calibration and (iv) the single camera-plane distance acquisition method.

Non-visual sensors use sound waves, infrared, pressure, electromagnetic induction etc. to sense the proximity of external targets to the robot and resolve the depth information of the target. The acquisition of target coordinates in the 3D space requires visual imaging and sensor completion, which increases the amount of data to be processed and the memory overhead of the robot. As a result, the reaction speed of the robot will be slow.

The deep-learning-based single-image depth estimation methods perform learning through neural networks, Markov random field etc. These methods estimate the depth of each small pixel in a single image and employ depth maps to represent the estimation results. These estimation results are generally not accurate enough for target localisation purpose. The accuracy of binocular perception distance is influenced by the performance of camera, illumination condition and baseline length. Due to the high complexity of binocular perception depth and the relatively large amount of data to be processed, the method is not able to satisfy real-time performance.

Traditional calibration methods could have high accuracy. In the calibration process, due to the limitation of equipment, it is still impossible to accurately record the coordinate of a point in the world coordinate system and the corresponding coordinate in the image labelling system accurately. If the coordinates are not accurately recorded, then the accuracy of coordinate transformation will also be affected accordingly. On the contrary, the self-calibration method does not rely on the calibration reference. However, the result of the self-calibration method is not so stable, comparing to traditional camera-based calibration methods.

The monocular plus plane mirror method, in which the plane mirror is always placed on the upper or sidewall of the tracking target, is applicable only when the tracking target is fixed in specific areas, such as fish in the tank. Although this method can effectively solve the matching and occlusion tracking problem of the target in the specific area, it is not a practical method since robots and targets are typically not fixed in a specific area.

To overcome the limitations of the existing methods, Mao *et al.* [43] studied the camera's imaging principle and the mathematical relationship between the target digitised length and distance, as well as, the effect of the change of the field of view (FOV) on the digitised length of the target during the imaging process. However, this work only studied the target distance measurement at the top of the target and at the centre of the vision.

In this paper, we propose a monocular geometric depth measurement method, which can well address the limitations (such as high computational complexities, high requirements for hardware and high cost) of traditional binocular camera depth perception algorithms. Further, the proposed method can overcome the limitations of the monocular camera calibration methods, which generally have high operational complexities, susceptible to coordinate points, and the camera needs to be recalibrated upon position movements as well as parameter changes. Additionally, our proposed method can solve the problem of the narrow application of the monocular camera-plane mirror methods.

The rest of this paper is organised as follows. Section 2 describes the method of target distance measurement when the top of the target is in the centre of vision. Then, Section 3 describes the method of distance measurement when the target is in the centre. Experimental results are presented in Section 4. Finally, conclusions are drawn in Section 5.

2 Target distance measurement when the top of target is in the vision centre

To solve the distance estimation problem based on monocular vision, we need to solve the problem when the top of the target is in the visual centre. Thus, we shall first introduce the principle of target imaging. We establish the space coordinate system, which is shown in Fig. 1, the height of the robot from the ground is denoted by h_1, and the height of the object AB (the red line AB in the figure) that the robot recognises is denoted by h_2. The target is located in front of the camera. The box labelled with 'CCD' represents the monocular camera of the robot head. In addition, the object AB is imaged on the CCD through the camera's pinhole. The goal is to find the distance of the target from the robot by using known parameters. In other words, we shall solve the problem of target distance [43]

$$|C_1D_1| = \frac{\sin\alpha}{2(1-\cos\alpha)}\frac{mh_2\sqrt{M^2+N^2}}{m^2+h_1^2-h_1h_2} \tag{1}$$

In order to measure the horizontal distance of the object AB from the camera, we assume the horizontal distance to be m, i.e. the length of line OA in Fig. 1. The robot is facing the top of the object AB, i.e. the focus of the camera in Fig. 1 is facing to point B. In our study, the bottom of the robot, i.e. the point O, is considered to be the centre of the coordinate system. The horizontal line along the vision direction is considered to be the X-axis, i.e. the line OA in Fig. 1. The line on the ground plane perpendicular to the OX axis is considered to be the Y-axis. Denoting the position of the camera as the point F and regarding the line OF as the Z-axis, the coordinate system shown in Fig. 1 can be established. Denoting the length of the line FC (i.e. the camera's focal length) as f and the image plane as π_1, the image of the object AB on the plane π_1 is CD after AB is imaged by the camera. It is evident that the points A, B, C, D, F and O are all in the plane XOZ. The relationship can be derived between |CD|, the focal length f, horizontal distance m, robot height h_1 and target height h_2 [43].

In (1), α is the FOV. Based on this, we can establish the relationship between the imaging length $|C_1D_1|$ of the target AB and the distance m of the target AB to the camera. The length $|C_1D_1|$ can be obtained through image segmentation. Finally, the distance m

© The Institution of Engineering and Technology 2020

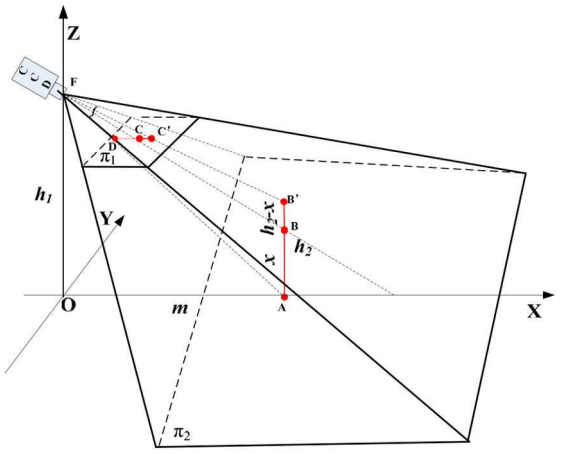

Fig. 2 *Illustration of camera imaging when the top of the target is above the centre position*

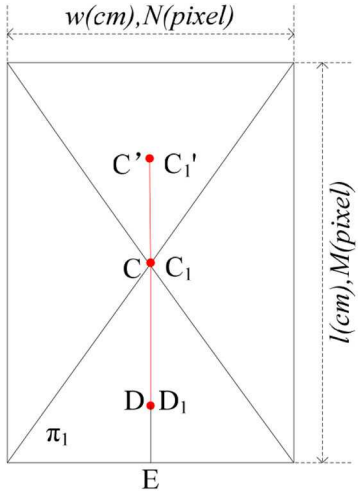

Fig. 3 *Target image plane π_1 when the top of the target is above the centre position*

can be calculated using the obtained relationship, which can be expressed as

(see (2)) .

In (2), the variable α can be calculated by the equivalent focal length. Then the target depth can be calculated from the digitalised image. In (2), the variables M and N depend on the camera resolution. The values of h_1 and h_2 are known beforehand.

By now, we have finished analysing target distance using monocular cameras when the target top is in the vision centre. In the next section, we shall study the problem of distance measurement when the target top is not in the vision centre.

3 Distance measurement when the target is in the centre of the vision

When the target top is not in the centre of the image, two possible cases may occur. The first case is that the target top is higher than the centre of the image, while the second case is that the target top is below the middle of the image.

3.1 Method of distance measurement when the top of the target is higher than the centre of the image

Fig. 2 shows the camera imaging when the target top is higher than the centre of the image, in which B' is the target top and B is the middle point of the camera. In addition, point C is the middle point of the image plane π_1. As we have mentioned in Section 2, the target distance can be calculated if the length $|AB|$ is known. Fig. 3 shows the image plane of π_1 of the target after the target is imaged. Assuming that the image of the target top B′ is C′ and $|AB| = x$, and, we obtain $|BB'| = h_2 - x$. So, the relationship between $|CD|$ and $|CC'|$ can be derived as

$$\frac{|CD|}{|CC'|} = \frac{x\left(m^2 + (h_1 - x)(h_1 - h_2)\right)}{h_2 - x\left(m^2 + h_1(h_1 - x)\right)} \tag{3}$$

In Fig. 3, $|C'D|$ is the target analogue image and $|C'_1D_1|$ is the target digitised image. It is evident that

$$\frac{|CD|}{|CC'|} = \frac{|C_1D_1|}{|C_1C'_1|} \tag{4}$$

Since both $|C_1D_1|$ and $|C_1C'_1|$ can be obtained by the target segmentation method, $|C_1D_1|/|C_1C'_1|$ is a fixed value. Let $|C_1D_1|/|C_1C'_1| = a$ and x be the distance from the image centre to the bottom of the target. According to (2), we can obtain the following equation: (see (5)) . There are two equations and two variables in (5). Theoretically, we can calculate the value of the target distance m. However, it is very difficult to find the solution of (5), which motivates us to find the approximated solution of (5). If the target is far from the camera, i.e. $m \gg h_1$, or the camera is significantly higher than the target, i.e. $h_1 \gg h_2$, we have

$$x \simeq \frac{ah_2}{1 + a} \tag{6}$$

If the target is closer and $m^2 \ll (h_1 - x)(h_1 - h_2)$, then we have

$$x \simeq \frac{ah_1h_2}{(h_1 - h_2) + ah_1} \tag{7}$$

$$m = \frac{h_2\sqrt{M^2 + N^2}\sin\alpha \; + \; \sqrt{h_2^2(M^2 + N^2)(\sin\alpha)^2 - 8|C_1D_1|(1 - \cos\alpha)\left(h_1^2 - h_1h_2\right)}}{4|C_1D_1|(1 - \cos\alpha)} \tag{2}$$

$$\begin{cases} a = \dfrac{x}{h_2 - x} \times \dfrac{m^2 + (h_1 - x)(h_1 - h_2)}{m^2 + h_1(h_1 - x)} \\ m = \dfrac{x\sqrt{M^2 + N^2}\sin\alpha \; + \; \sqrt{x^2(M^2 + N^2)(\sin\alpha)^2 - 16|C_1D_1|(1 - \cos\alpha)\left(h_1^2 - h_1x\right)}}{4|C_1D_1|(1 - \cos\alpha)} \end{cases} \tag{5}$$

© The Institution of Engineering and Technology 2020

By substituting (6) or (7) into the second equation in (5), we can finally solve the distance of the target.

3.2 Method of distance measurement when the top of the target is below the centre position

When the top of the target is below the centre position, we denote the intersection between the extension of the target AB and visual centre line as B′, we assume the imaging points of B and B′ on the image plane π_1 to be C and C′, respectively. It is easy to find that C′ is the central point of the image plane π_1. The image plane is shown in Fig. 4. As mentioned in Section 2, we can find the solution of m if we know the value of $|AB'|$. After the target is imaged, the target image plane π_1 forms as shown in Fig. 4. Denoting $|AB'| = x$, we have $|BB'| = x - h_2$. As in the previous section, we can easily find the following relationship: (see (8)) . In (8), D_1, C_1 and C'_1, which are the digitalised points of D, C and C′, can be obtained by using the technique of image segmentation. Therefore, the variable a in (8) is constant.

Since it is difficult to get the value of m, we can use an approximation method to find an approximated solution of (8). When $m \gg h_1$ or $h_1 \gg h_2$, we have

$$x \simeq \frac{ah_2}{a-1} \tag{9}$$

When $m^2 \ll (h_1 - x)(h_1 - h_2)$, we have

$$x \simeq \frac{ah_1h_2}{h_1(a-1)+h_2} \tag{10}$$

Then, by substituting (9) or (10) into the second equation of (8), we can obtain the target depth in this case.

From (2), (5), (8), we can find that in order to measure the target distance, we must segment the target bottom D_1, target vertex C_1, and image centre point C'_1 from the image. After that, the target distance can be calculated according to the formula.

4 Experimental results

4.1 Experimental settings

We use NIKON D7100 in our study. NIKON D7100 is a SLR camera launched by NIKON in 2013. The conversion ratio of the focal length between the camera and the 135-specification camera is 1.5 times rate. This means that if the focal length of D7100 is 18 mm, the equivalent focal length of the 135-specification camera should be 27 mm to achieve the same FOV. By using the relationship between the focal length and the FOV, we can calculate the α of the FOV of the camera with any value of the focal length. We use a camera with a fixed focal length of 50 mm and a CCD size of 23.5mm × 15.6 mm, so the FOV is 0.549838768273275°.

The experiments are performed in the environment of Visual Studio2013. To take photos of targets with different distances and heights, we have designed a simple mobile shooting platform that can be raised and lowered. Meanwhile, the target we choose is green and columnar with fixed height (69 cm). We adopt the colour-based image segmentation method to segment the target object [44].

4.2 Experimental results and analysis

To validate the efficiency of our proposed method, we conduct experiments with different distances from the target to the camera, different heights of the camera and different environments. First, we carry out the experiment by setting the target depths to be 150, 200 and 300 cm, respectively, and the heights of the camera to be 81 and 71 cm, respectively. Experimental environment and segmentation of target are shown in Fig. 5. Five images are taken for each case. The target distance is calculated using our proposed method. The results are shown in Table 1. To evaluate the performance of our algorithm, we define the concept of deviation error as

$$P_e = \frac{|m_{tad} - m_{tcd}|}{m_{tad}} \times 100\% \tag{11}$$

In (11), m_{tad} denotes the actual distance of the target and m_{tcd} denotes the target distance obtained using our proposed algorithm.

Our results in Table 1 show that the average errors are 2.566, 1.586 and 0.829% when the distances from the object to the camera to be 150, 200 and 250 cm, respectively. The average errors are 0.969 and 1.812% when the heights of the camera are 81 and 71 cm, respectively. The total average error rate is 1.660% with mean square error is 1.232%. Based on these results, it is clear that our proposed method is robust. The overall average error is 1.343%, while the largest and smallest errors are 3.708 and 0.066% respectively, which clearly verifies the effectiveness of our proposed method.

To validate the method proposed further, we conducted another experiment to test the amount of data. We conducted six sets of experiments with the distances from target to the camera to be 200, 250 and 300 cm, respectively. The camera heights are 81 and 88 cm, respectively, and ten images are taken for each experiment. The experimental results are shown in Fig. 6.

Fig. 6 shows the result of actual depth and measured depth using our method. Statistical analysis has also been performed for the results above. The results show that the average errors are 2.442, 1.532 and 1.399% when the distances from target to the camera are 200, 250 and 300 cm respectively, while the average errors are 1.489 and 2.093% when the heights of the camera are 88 and 81 cm, respectively. The overall average error is 1.791 cm. The statistical analysis further illustrates the rationality of the approximated solution of (5) and (8).

To further evaluate the feasibility of our proposed algorithm, additional object (i.e. a 50 cm steel pipe) has now been used as the experimental target. To facilitate the segmentation, we dye the steel pipe to be red, as shown in Fig. 7. Six sets of experiments have

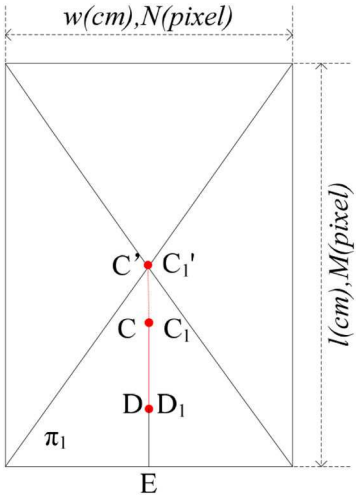

Fig. 4 *Target image plane π_1 when the top of the target is below the centre position*

$$\begin{cases} a = \dfrac{|DC'|}{|CC'|} = \dfrac{|D_1C'_1|}{|C_1C'_1|} = \dfrac{x}{x-h_2} * \dfrac{m^2 + (h_1 - x)(h_1 - h_2)}{m^2 + h_1(h_1 - x)} \\ m = \dfrac{x\sqrt{M^2+N^2}\sin\alpha + \sqrt{x^2(M^2+N^2)(\sin\alpha)^2 - 16|C_1D_1|(1-\cos\alpha)(h_1^2 - h_1x)}}{4|C_1D_1|(1-\cos\alpha)} \end{cases} \tag{8}$$

© The Institution of Engineering and Technology 2020

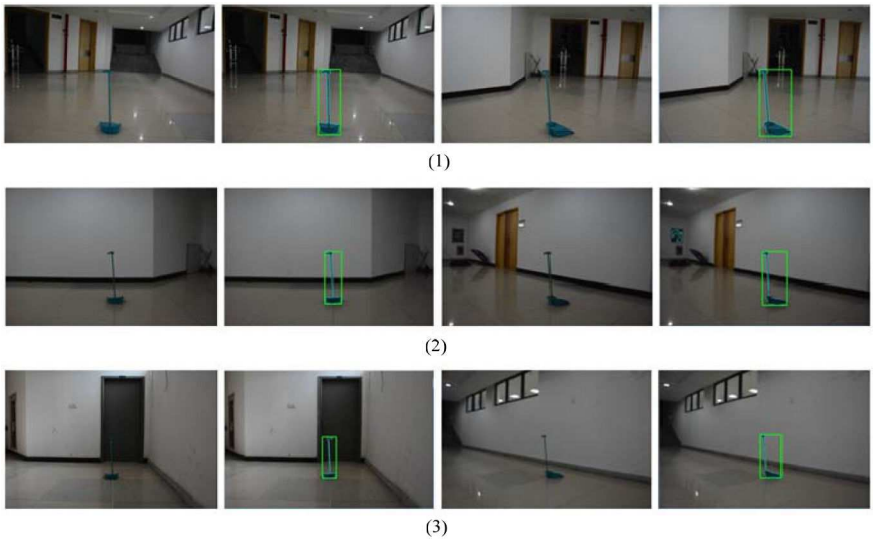

Fig. 5 *Illustration of the experimental environment and segmentation of target. Target depths are set to 150, 200 and 250 cm in (1)–(3), respectively*

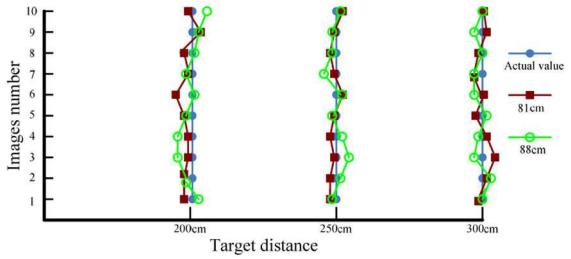

Fig. 6 *Illustration of actual distance and the measured distance using our method*

Fig. 7 *Target of 50 cm red steel pipe*

been carried out with the target distances of 583, 675, 743 cm and camera heights of 83, 107.5 cm, respectively. Five images are taken for each set of experiment. We then employ the proposed method to calculate the target distance. The experimental results are shown in Table 2. The average errors in the table are calculated using (11).

From Table 2, we can find that our proposed method achieves results with a maximum and minimum error of 3.290 and 0.031%, respectively, and an average error <2% for all sets of experiments. The results could thus further establish the feasibility of our method to calculate the target distance.

4.3 Discussion

Table 3 lists the differences between our depth measurement methods and related methods. Based on the results, we can conclude that the exiting depth measurement methods mainly include target depth [10, 12, 36] and depth mapping [13, 14, 15, 18]. In [10], the monocular camera and azimuth prediction are used to measure indoor targets. This method requires other objects in the room as reference objects to predict the azimuth, and the target depth deviation is large with an average deviation of 0.72 cm. In [12], two radar CCDs were used to measure the depth of the airborne target, with an average deviation of 2.8 cm. Neither of these methods is suitable for robot vision.

Doh *et al.* [13] employed two SLAM algorithms to deal with a semi-permanent dynamic induced by the door opening and closing. Yang *et al.* [14] aim to estimate the motion of the robot and

Table 1 Experimental results of the actual distance of the target and the measured depth by employing our method

TAD, cm	CH, cm	IN	TCD, cm	P_e %%
150	81	1	152:318	1:545
		2	148:351	1:099
		3	152:318	1:545
		4	147:972	1:352
		5	147:672	1:552
	71	1	154:258	2:838
		2	154:443	2:962
		3	154:443	2:962
		4	155:602	3:708
		5	155:460	3:64
200	81	1	198:497	0:751
		2	198:497	0:751
		3	199:869	0:066
		4	199:869	0:066
		5	201:255	0:628
	71	1	193:151	2:425
		2	198:602	0:699
		3	197:438	1:281
		4	197:438	1:281
		5	202:183	1:091
250	81	1	248:128	0:749
		2	248:626	0:549
		3	248:414	0:634
		4	251:628	0:651
		5	251:966	0:786
	71	1	247:723	0:911
		2	247:892	0:843
		3	246:840	1:264
		4	247:809	0:876
		5	248:005	0:798

TAD is the target actual distance; CH is the camera height, IN is the images number and TCD is the target calculation distance.

© The Institution of Engineering and Technology 2020

Table 2 Experimental results on the target of red steel pipe

TAD, cm	CH, cm	IN	TCD, cm	P_e %
583	83	1	563.812	3.290
		2	565.910	2.931
		3	583.183	0.031
		4	565.313	3.033
		5	590.247	1.243
		average	573.494	1.631
	107.5	1	583.244	0.042
		2	573.418	1.644
		3	577.129	1.007
		4	578.547	0.764
		5	576.050	1.192
		average	577.678	0.913
675	83	1	656.456	2.747
		2	674.643	0.053
		3	657.912	2.532
		4	682.313	1.083
		5	673.244	0.230
		average	668.914	0.902
	107.5	1	675.764	0.113
		2	665.493	1.408
		3	669.825	0.767
		4	670.597	0.652
		5	668.950	0.896
		average	670.126	0.722
743	83	1	760.465	2.351
		2	755.146	1.635
		3	723.420	2.635
		4	739.548	0.465
		5	755.313	1.657
		average	746.777	0.508
	107.5	1	751.461	1.139
		2	739.548	0.465
		3	728.361	1.970
		4	739.101	0.525
		5	728.312	1.977
		average	737.358	0.759

Table 3 Results: quantitative comparison of various methods

Works	Methods	Object	Results	Accuracy
[10]	one CCD, SLAM	Indoor targ.	depth	0:71 m
[12]	two CCD	Aircraft	depth	2:8 cm
[13]	two SLAM	Indoor targ.	depth maps	unreported
[14]	RGB-D camera, DRE-SLAM	Ground dynamic target	pose translation	RMSE: 0.0187 m
[15]	laser, CCD, SLAM	Indoor	topological map	unreported
[16]	PP-MRF	A sing. image	depth maps	71:2%
[17]	FCNN	A sing. image	depth maps	93%
[18]	DCNN	A sing. image	depth	unreported
[21]	HGR	An RGB image	depth maps	96:92%
[39]	one CCD, a mirror	fish	depth	96:925%
proposed work	one camera	Ground target	depth	98:657%

SLAM is the simultaneous localisation and mapping, PP-MRF is the plane parameters-Markov random field, FCNN is the fully convolutional neural network, CNN is the deep convolutional neural network, GM-ADP is the geometric model-analogue to digital principle, HGR is the hierarchical guidance and regularisation network and RMSE is the root mean square error.

construct a static background OctoMap in both dynamic and static environments. The root mean square error of pose translation is 0.0187 m. Fernandez *et al.* [15] utilised the internal odometry and computed visual odometry to build the metric map.

The literature works proposed in [16–18, 21] aim for depth estimation of static images, and the results are all depth mapping (depth information of image pixel blocks). These methods require deep learning with high computational complexity. They are suitable for computer vision but not machine vision as well as real-time target tracking for the robot. In [36], the single-camera plus flat mirror method is used to measure the target depth in the water. This method can only be performed in a close-up environmen, and is not suitable for tracking of the target by moving robot.

In our experiment, due to the simple experimental equipment and the lack of precision instruments, there are certain deviations in focusing, distance measurement and height measurement. In

© The Institution of Engineering and Technology 2020

addition, although we use green props as the target for the ease of segmentation, we can hardly segment the target with 100% accuracy due to the deficiency of the current method for image segmentation.

Thus, the errors mentioned above are in the normal deviation range. These errors can further be reduced by increasing the accuracy of instruments.

5 Conclusions

In this work, we propose a novel method of measuring target distance according to imaging principle and the principle of analogy signal transforming to the digital signal. We theoretically derive the relationship between the target distance, FOV, equivalent focal length and image resolution. Our method can effectively measure the target depth using a single camera. Experimental results verify that our method is robust in both theory and practices. Moreover, the proposed method is simple and can effectively reduce production cost. Therefore, our method can be implanted into mobile phones and webcams to avoid the complicated stereo matching when measuring target distance using binocular cameras. Additionally, our method is efficient and can satisfy the real-time requirements in industrial robotic productions.

6 Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61771430, 61573316 and 61873082), the Zhejiang province Natural Science Foundation of China (No. LY20F020022) and the National Key R&D Program of China (No. 2018YFB0204003).

7 References

[1] Wang, H., Zhao, J., Zhao, J.W., *et al.*: 'A new rapid-precision position measurement method for a linear motor mover based on a 1-D EPCA', *IEEE Trans. Ind. Electron.*, 2018, **65**, (9), pp. 7485–7494

[2] Hachmon, G., Mamet, N., Sasson, S., *et al.*: 'A non-newtonian fluid robot', *Artif. Life*, 2016, **22**, (1), pp. 1–22

[3] Tan, K.H.: 'Squirrel-cage induction generator system using wavelet petri fuzzy neural network control for wind power applications', *IEEE Trans. Power Electron.*, 2016, **31**, (7), pp. 5242–5254

[4] Kong, L.F., Wu, P.L., Li, X.s.: 'Object depth estimation using translations of hand eye system with uncalibrated camera', *Comput. Integr. Manuf. Syst.*, 2009, **18**, (5), pp. 1633–1639

[5] Hoang, N.B., Kang, H.J.: 'Neural network-based adaptive tracking control of mobile robots in the presence of wheel slip and external disturbance force', *Neurocomputing*, 2016, **18**, (5), pp. 12–22

[6] Mendes, N., Neto, P.: 'Indirect adaptive fuzzy control for industrial robots: a solution for contact applications', *Expert Syst. Appl.*, 2015, **42**, (22), pp. 8929–8935

[7] Ghommam, J., Mehrjerdi, H., Saad, M.: 'Robust formation control without velocity measurement of the leader robot', *Control Eng. Pract.*, 2013, **21**, (8), pp. 1143–1156

[8] Charalampous, K., Kostavelis, I., Gasteratos, A.: 'Thorough robot navigation based on SVM local planning', *Robot. Auton. Syst.*, 2015, **70**, (8), pp. 166–180

[9] Jia, T., Shi, Y., Zhou, Z.X., *et al.*: '3D depth information extraction with omni-directional camera', *Inf. Process. Lett.*, 2015, **115**, (2), pp. 285–291

[10] Lee, T.J., Kim, C.H., Cho, D.I.D.: 'A monocular vision sensor-based efficient SLAM method for indoor service robots', *IEEE Trans. Ind. Electron.*, 2019, **66**, (1), pp. 318–328

[11] Steinvall, O.: 'Effects of target shape and reflection on laser radar cross sections', *Appl. Opt.*, 2000, **39**, (24), pp. 4381–4391

[12] Yao, J.: 'Image registration and superposition for improving ranging accuracy of imaging Laser radar', *Chin. J. Lasers*, 2010, **37**, (6), pp. 1613–1617

[13] Doh, N.L., Lee, K., Chung, W.K., *et al.*: 'Simultaneous localization and mapping algorithm for topological maps with dynamics', *IET Control Theory Appl.*, 2009, **3**, (9), pp. 1249–1260

[14] Yang, D.S., Bi, S.S., Wang, W., *et al.*: 'DRE-SLAM: dynamic RGB-D encoder SLAM for a differential-drive robot', *Remote Sens.*, 2019, **11**, (4), pp. 1–29

[15] Fernandez, L., Paya, L., Reinoso, O., *et al.*: 'Appearance-based approach to hybrid metric-topological simultaneous localization and mapping', *IET Intell. Transp. Syst.*, 2014, **8**, (8), pp. 688–699

[16] Saxena, A., Sun, M., Nang, A.Y.: 'Make3d: learning 3D scene structure from a single still image', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (5), pp. 824–840

[17] Haim, H., Elmalem, S., Giryes, R., *et al.*: 'Depth estimation from a single image using deep learned phases coded mask', *IEEE Trans. Comput. Imag.*, 2018, **4**, (3), pp. 298–310

[18] Liu, F., Shen, C., Lin, G., *et al.*: 'Learning depth from single monocular images using deep convolutional neural fields', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, **38**, (10), pp. 2024–2039

[19] Eigen, D., Puhrsch, C., Fergus, R.: 'Depth map prediction from a single image using a multi-scale deep network'. 28th Annual Conf. on Neural Information Processing Systems, Montreal, QC, Canada, December 2014, pp. 2366–2374

[20] He, L., Wang, G.H., Hu, Z.Y.: 'Learning depth from single images with deep neural network embedding focal length', *IEEE Trans. Image Process.*, 2018, **27**, (9), pp. 4676–4689

[21] Zhang, Z.Y., Xu, C.Y., Yang, J., *et al.*: 'Deep hierarchical guidance and regularization learning for end-to-end depth estimation', *Pattern Recognit.*, 2018, **83**, (11), pp. 430–442

[22] Li, H, Zhang, X.M., Zeng, L., *et al.*: 'A monocular vision system for online pose measurement of a 3RRR planar parallel manipulator', *J. Intell. Robot. Syst.*, 2018, **92**, (1), pp. 3–17

[23] Wardle, S.G., Palmisano, S., Gillam, B.J.: 'Monocular and binocular edges enhance the perception of stereoscopic slant', *Vis. Res.*, 2014, **100**, pp. 113–123

[24] Yang, J., Xu, R., Ding, Z., *et al.*: '3D character recognition using binocular camera for medical assist', *Neurocomputing*, 2017, **220**, pp. 17–22

[25] Zhu, S.P, Gao, Y.: 'Nonconstant 3-d coordinate measurement of cross-cutting feature points on the surface of a large-scale workpiece based on the machine vision method', *IEEE Trans. Instrum. Meas.*, 2017, **59**, (7), pp. 1874–1887

[26] Li, J., Allinson, N.M.: 'A comprehensive review of current local features for computer vision', *Neurocomputing*, 2008, **71**, (10-12), pp. 1771–1787

[27] Song, L.M., Wu, W.F., Guo, J.R., *et al.*: 'Survey on camera calibration technique'. 5th Int. Conf. on Intelligent Human-Machine Systems and Cybernetics (IHMSC 2013), Hangzhou, Zhejiang, China, August 2013, pp. 389–392

[28] Chen, W.P., Wu, M.Y., Chen, J., *et al.*: 'Research on autocalibration technology of intelligent vehicle camera based on machine vision'. The 2nd Int. Conf. on Image, Vision and Computing, Chengdu, China, July 2017, pp. 684–687

[29] Sun, J., Gu, H.B.: 'Research of linear camera calibration based on planar pattern', *World. Acad. Sci. Eng. Technol.*, 2009, **36**, pp. 628–632

[30] Carlos, R.V., Antonio-Jose, S.S.: 'Optimal conditions for camera calibration using a planar template'. The 18th IEEE Int. Conf. on Image Processing, Brussels, Belgium, September 2011, pp. 853–856

[31] Yang, X.F., Huang, Y.M., Gao, F.: 'A simple camera calibration method based on sub-pixel corner extraction of the chessboard image'. Proc. - 2010 IEEE Int. Conf. on Intelligent Computing and Intelligent Systems (ICIS 2010), Xiamen, China, October 2010, pp. 688–692

[32] Park, J.H., Park, S.H.: 'Improvement on Zhang's camera calibration', *Appl. Mech. Mater.*, 2014, **479-480**, pp. 170–173

[33] Abdel-Aziz, Y.I., Karara, H.M.: 'Direct linear transformation into object space coordinates in close-range photogrammetry', *Photogrammetric Engineering and Remote Sensing*, 2015, **81**, (2), pp. 103–107

[34] Tsai, : 'A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses', *IEEE J. Robot. Autom.*, 1987, **RA-3**, (4), pp. 323–344

[35] Zhang, Z.: 'Flexible camera calibration by viewing a plane from unknown orientations'. Proc. Seventh Int. Conf. on Computer Vision, Kerkyra, Greece, September 1999, pp. 666–673

[36] Zhang, Z.Y.: 'Camera calibration with one-dimensional objects', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (7), pp. 892–899

[37] Pollefeys, M., Koch, R., Van, L.G.: 'Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters'. Proc. of the 6th Int. Conf. on Computer Vision, Bombay, India, September 1998, pp. 90–95

[38] Ferran, E.: 'A new linear method for camera self-calibration with planar motion', *Math. Imag. Vis.*, 2007, **27**, (1), pp. 81–88

[39] Mao, J., Xiao, G., Sheng, W., *et al.*: 'Research on realizing the 3D occlusion tracking location method of fish's school target', *Neurocomputing*, 2016, **214**, pp. 61–79

[40] Wang, Q., Fu, L., Liu, Z.Z.: 'Review on camera calibration'. 2010 Chinese Control and Decision Conf. (CCDC 2010), Xuzhou, China, May 2010, pp. 3354–3358

[41] Li, S.Q., Xie, X.P., Zhuang, Y.J.: 'Research on the calibration technology of an underwater camera based on equivalent focal length', *Meas., J. Int. Meas. Confederation*, 2018, **122**, pp. 275–283

[42] Laurel, B.J., Laurel, C.J., Brown, J.A., *et al.*: 'A new technique to gather 3-D spatial information using a single camera', *J. Fish Biol.*, 2005, **66**, pp. 429–441

[43] Mao, J.F., Zhang, M.G., Zhu, L., *et al.*: 'Target depth measurement for machine monocular vision'. The Pacific-Rim Conf. on Multimedia (PCM) 2017, Harbin, China, September 2017, pp. 18–29

[44] Yue, X.D., Miao, D.Q., Zhang, N., *et al.*: 'Multiscale roughness measure for color image segmentation', *Inf. Sci.*, 2012, **216**, pp. 93–112

© The Institution of Engineering and Technology 2020