Proceedings of the 2016 IEEE
International Conference on Robotics and Biomimetics
Qingdao, China, December 3-7, 2016

A Monocular Vision System for Pose Measurement in Indoor Environment

Lingyi Xu, Zhiqiang Cao, *Senior Member, IEEE*, Xilong Liu

Abstract—**This paper proposes a new localization method using monocular camera. A camera is hanging on the ceiling toward the ground. Its optical axis is perpendicular to the ground as possible. Its intrinsic parameters are calibrated in advance offline. Its extrinsic parameters refer to the chessboard reference are calibrated online. Then a method is proposed to estimate the 3D positions and orientations of the specified arrow targets in the reference frame with only one captured image. The experimental results in indoor environment verify the effectiveness of the proposed method.**

Key words: **Visual measurement, localization, position estimation, monocular vision, mobile robot**

I. Introduction

Camera is a typical sensor used in mobile robots for localization and navigation, which is also a viewer friendly. Most mobile robot systems are equipped a vision system. Vision systems are classified to two categories depending on the number of cameras: monocular vision system and multi-camera vision system. Monocular vision system is more common used since its space and energy efficient [1][2]. One kind of monocular vision system consists of an omnidirectional camera. Omnidirectional camera has large view field but with tremendous distortion. Goedeme *et al.* installed an omnidirectional camera on the top of robot and used Bayesian approach to estimate its position [3]. In [4], a localization method combining maximum likelihood estimation and Gaussian process regression was proposed. The more popular kind of monocular vision system consists of a traditionally normal camera. Compared to the approach with omnidirectional cameras, localization model using normal cameras is computational and timing efficiency. Mostegel *et al.* proposed a measurement method of estimating the influence of possible camera motions on Micro Aerial Vehicles (MAVs). The image sequence is needed to estimate the camera's position [5][6]. Yin *et al.* used PnP method to calculate the position of points only on the ground [7]. In [8], a normal monocular camera upward to the ceiling was installed on the top of mobile robot. The natural cross points on the ceiling were used to locate the robot on the condition of flat floor. In [9], a vertical reference chessboard was placed in front of the camera, perpendicular to its optical axis. It's able to estimate the height of target in the same plane with reference chessboard. In the monocular vision systems above [5-9], the objects to be measured are limited to a single plane, and the small field of view can only provide local and little information of the environment. All these shortages limit their applications.

Monocular camera fixed on ceiling and downward to the ground gives a large and global view of the environment. It monitors the targets in whole vision field from above [10-12]. In [10], a colorful landmark was used to localize the robot in the field. A 3-dimensional (3D) model known metal bracket was tracked in [11]. The small movement of bracket was realized by hand. Those methods require the condition that the optical axis of camera is perpendicular to the ground, which is hardly to be ensured. Besides, the height of targets may be different. In this way, they are not suitable.

In this paper, a novel monocular method is presented to estimate 3D positions and orientations for targets on the ground with the camera fixed on the ceiling. It firstly estimates the position and orientation of the world frame defining by the chessboard relative to the camera frame. Then the orientation of the world frame relative to the camera frame is employed to estimate the points' positions or the specified arrow target's pose in the world frame. Compared to typical PnP methods, this method can estimate the orientation angle and 3D positions of the specified arrow target with a single image. It can also estimate 2D positions on the plane defining by the chessboard.

The rest of this paper is organized as follows. In Section II, the localization model is presented. Section III gives the calibration principle and method. Section IV shows the measurement method. Experiments are provided in Section V. Section VI concludes the paper.

II. Localization Model

For the ceiling-fixed monocular visual measurement system, the optical axis of camera isn't always perpendicular to the ground, which means the approximately vertical model isn't suitable for this condition. A modified model is shown in Fig. 1. The camera is fixed on the ceiling, looking down from above. The origin O_c of the camera frame is at the optical center. The orientation of axis X_c is corresponding to the horizontal axis U in the image. The direction of axis Y_c is corresponding to the vertical axis V in the image. The optical axis direction to the scene is selected as the axis Z_c. The

This work was supported in part by the National Natural Science Foundation of China under Grant 61273352.

L. Xu is with the Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, New Brunswick, NJ 08854-8058, USA (e-mail: lingyixu@yahoo.com).

Z. Cao and X. Liu are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

origin O_c shares the same position with O'_w, as the origin of the medium frame $O'_wX'_wY'_wZ'_w$.The axis X'_w is paralleled with the axis X_w sharing the same direction, which is assigned by needed. The direction of axis Z'_w is paralleled with Z_w heading to ground along the line perpendicular to ground. The axis Y'_w is assigned by Right-hand rule. The coordinates $O_wX_wY_wZ_w$ is the world frame.

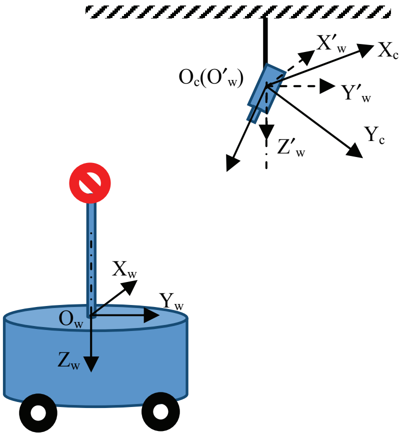

Fig. 1 Monocular vision system and coordinates

The extrinsic parameters matrix of a camera indicates the transformation from the camera's frame to the world frame. It can be written as (1),

$$ {}^{c}T_{w} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1} $$

where $[n_x\ n_y\ n_z]^T$, $[o_x\ o_y\ o_z]^T$, and $[a_x\ a_y\ a_z]^T$ represent the unit vectors of the X_w-, Y_w- and Z_w-axis of the world frame expressed in the camera frame, $[p_x\ p_y\ p_z]^T$ is the position vector of the origin of the world frame described in the camera frame.

The coordinates for any point in the world frame is able to be transformed into the camera frame with the extrinsic parameters matrix.

$$ \begin{bmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \\ 1 \end{bmatrix} = \begin{bmatrix} z_{ci}x_{1ci} \\ z_{ci}y_{1ci} \\ z_{ci} \\ 1 \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \tag{2} $$

where (x_{ci}, y_{ci}, z_{ci}) are the Cartesian coordinates of the i-th point in the camera frame, (x_{wi}, y_{wi}, z_{wi}) are its coordinates in the world frame, (x_{1ci}, y_{1ci}) are the normalized imaging coordinates.

The normalized imaging coordinates (x_{1ci}, y_{1ci}) are computed from the image coordinates and the intrinsic parameters. The imaging model of camera is described by the intrinsic parameters, as given in (3).

$$ \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1ci} \\ y_{1ci} \\ 1 \end{bmatrix} \tag{3} $$

where k_x, k_y are the magnification factors along X- and Y-axis, (u, v) are the image coordinates of the feature point. (u_0, v_0) are the image coordinates of primary point on imaging plane.

Assuming z_{wi}=0, formula (2) expands into (4). p_z always be positive in this model, which means all elements in (4) could be divided by p_z. In this way, (5) is satisfied for each one of the points on the normalized plane of camera.

$$ \begin{cases} \left(n_z x_{wi} + o_z y_{wi} + p_z\right) x_{1ci} = n_x x_{wi} + o_x y_{wi} + p_x \\ \left(n_z x_{wi} + o_z y_{wi} + p_z\right) y_{1ci} = n_y x_{wi} + o_y y_{wi} + p_y \end{cases} \tag{4} $$

$$ \begin{cases} n'_x x_{wi} - n'_z x_{wi} x_{1ci} + o'_x y_{wi} - o'_z y_{wi} x_{1ci} + p'_x = x_{1ci} \\ n'_y x_{wi} - n'_z x_{wi} y_{1ci} + o'_y y_{wi} - o'_z y_{wi} y_{1ci} + p'_y = y_{1ci} \end{cases} \tag{5} $$

where $n'_x=n_x/p_z$, $n'_y=n_y/p_z$, $n'_z=n_z/p_z$, $o'_x=o_x/p_z$, $o'_y=o_y/p_z$, $o'_z=o_z/p_z$, $p'_x=p_x/p_z$, $p'_y=p_y/p_z$.

Each feature point has two equations in the form of (5). For n feature points, there would be $2n$ equations which could be written in matrix format like (6).

$$ AX = B \tag{6} $$

where

$$ A = \begin{bmatrix} x_{w1} & 0 & -x_{w1}x_{1c1} & y_{w1} & 0 & -y_{w1}x_{1c1} & 1 & 0 \\ 0 & x_{w1} & -x_{w1}y_{1c1} & 0 & y_{w1} & -y_{w1}y_{1c1} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{wn} & 0 & -x_{wn}x_{1cn} & y_{wn} & 0 & -y_{wn}x_{1cn} & 1 & 0 \\ 0 & x_{wn} & -x_{wn}y_{1cn} & 0 & y_{wn} & -y_{wn}y_{1cn} & 0 & 1 \end{bmatrix} $$

$$ B = \begin{bmatrix} x_{1c1} & y_{1c1} & \cdots & x_{1cn} & y_{1cn} \end{bmatrix}^T $$

$$ X = \begin{bmatrix} n'_x & n'_y & n'_z & o'_x & o'_y & o'_z & p'_x & p'_y \end{bmatrix}^T $$

The value of X is calculated by Least-Square Method (LS Method). Then the value of p_z is calculated from (7) as the prerequisite of n, o, p calculation.

$$ p_z = \frac{1}{\left\| n' \right\|} \tag{7} $$

$[n_x\ n_y\ n_z]^T$, $[o_x\ o_y\ o_z]^T$, and $[p_x\ p_y]^T$ are obtained from the multiplication of X and p_z. Then $[a_x\ a_y\ a_z]^T$ is calculated with $[n_x\ n_y\ n_z]^T \times [o_x\ o_y\ o_z]^T$.

III. PARAMETERS CALIBRATION

When trying to install the monocular camera on the ceiling, it's hardly to say the X_c axis is parallel to the ground or not. In this case, the transformation from the camera frame $O_cX_cY_cZ_c$ to the medium frame $O'_wX'_wY'_wZ'_w$ is (8).

$$
{}^{c}T_{w1} = \begin{bmatrix} n_x & o_x & a_x & 0 \\ n_y & o_y & a_y & 0 \\ n_z & o_z & a_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}
$$

Since there only exists translation between the world frame $O_wX_wY_wZ_w$ and the medium frame $O'_wX'_wY'_wZ'_w$, the transformation ${}^{w'}T_w$ is (9):

$$
{}^{w'}T_w = Trans(a,b,h) = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & h \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}
$$

where (a, b, h) means the position of the origin of world coordinate respects to the medium coordinate.

Combining the two transformations in (8) and (9) gives the expression of transformation from the world frame to the camera frame, as given in (10).

$$
{}^{c}T_w = \begin{bmatrix} n_x & o_x & a_x & an_x + bo_x + ha_x \\ n_y & o_y & a_y & an_y + bo_y + ha_y \\ n_z & o_z & a_z & an_z + bo_z + ha_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}
$$

The comparison of (1) and (10) gives the way to estimate the values of a, b, and h by (11), where $a_x=n_yo_z-n_zo_y$, $a_y=n_zo_x-n_xo_z$, $a_z=n_xo_y-n_yo_x$.

$$
\begin{cases} an_x + bo_x + ha_x = p_x \\ an_y + bo_y + ha_y = p_y \\ an_z + bo_z + ha_z = p_z \end{cases} \tag{11}
$$

Assuming there are 4 known points in image, which are used as references. The positions of those points represented in world frame are known. Once their image coordinates are extracted from image, their normalized imaging coordinates can be computed with (3). Then the extrinsic parameters can be estimated with (6) by LS method. Theoretically, 4 points are enough to calculate the extrinsic matrix while more points may give the higher accuracy and redundancy at same time. After obtain extrinsic matrix, (11) is used to estimate the parameters a, b and h.

IV. Measurement

A. Position Measurement for Points on Chessboard Plane

Once all parameters are calibrated, the position and orientation of targets in the vision field are easily estimated. Combining extrinsic matrix (2) and intrinsic matrix (3) gives (12).

$$
\begin{cases} (n_z x_{wi} + o_z y_{wi} + a_z z_{wi} + p_z) x_{1ci} \\ = n_x x_{wi} + o_x y_{wi} + a_x z_{wi} + p_x \\ (n_z x_{wi} + o_z y_{wi} + a_z z_{wi} + p_z) y_{1ci} \\ = n_y x_{wi} + o_y y_{wi} + a_y z_{wi} + p_y \end{cases} \tag{12}
$$

In (12), n, o, a, p are calculated from (6), (7) and (10). On the chessboard plane, z_{wi}=0. Submitting z_{wi}=0 to (12) and re-arranging it, we have

$$
\begin{cases} (n_z x_{1ci} - n_x) x_{wi} + (o_z x_{1ci} - o_x) y_{wi} = p_x - p_z x_{1ci} \\ (n_z y_{1ci} - n_y) x_{wi} + (o_z y_{1ci} - o_y) y_{wi} = p_y - p_z y_{1ci} \end{cases} \tag{13}
$$

It can be found that the coordinates (x_{wi}, y_{wi}) of points on the chessboard plane in the world frame can be computed with (13) from a single image.

B. Pose Estimation for Specified Arrow Targets

There are five corner points in the specified arrow target. The target frame is established on the center of the rectangle formed with the four corners except the corner with sharp angle. The coordinates of the five corners in the target frame are known. The transformation ${}^{c}T_g$ from the camera frame to the target frame is decomposed to three transformations such as a rotation transformation ${}^{c}T_{w1}$ from the camera frame to the medium frame, a translation transformation ${}^{w1}T_{g1}$ from the medium frame to the target medium frame, and a rotation transformation ${}^{g1}T_g$ from the target medium frame to the target frame. The target medium frame has same origin as the target frame and has same orientations as the world frame. ${}^{w1}T_{g1}$ and ${}^{g1}T_g$ are given in (14) and (15). ${}^{c}T_g$ is given in (16).

$$
{}^{w1}T_{g1} = \begin{bmatrix} 1 & 0 & 0 & a_g \\ 0 & 1 & 0 & b_g \\ 0 & 0 & 1 & h_g \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{14}
$$

$$
{}^{g1}T_g = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{15}
$$

Similar to (4), formula (17) is derived from (16). It is sure that h_g>0. After divided by h_g, (17) is rewritten to (18). There are four unknown variables $\cos\theta/h_g$, $\sin\theta/h_g$, a_g/h_g, and b_g/h_g in (18). They can be solved via LS method with no less than two points on the target. Then the angle θ is known. Formula (18) is rewritten to (19) with the consideration of known $\cos\theta$ and $\sin\theta$ terms. Then a_g, b_g and h_g can be solved via LS method from (19).

The coordinates of original point of the target in the world frame are computed as

$$
{}^{c}T_{g}={}^{c}T_{w1}{}^{w1}T_{g1}{}^{g1}T_{g}=\begin{bmatrix} n_x & o_x & a_x & 0 \\ n_y & o_y & a_y & 0 \\ n_z & o_z & a_z & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & a_g \\ 0 & 1 & 0 & b_g \\ 0 & 0 & 1 & h_g \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

$$
=\begin{bmatrix} n_x\cos\theta+o_x\sin\theta & -n_x\sin\theta+o_x\cos\theta & a_x & a_g n_x+b_g o_x+h_g a_x \\ n_y\cos\theta+o_y\sin\theta & -n_y\sin\theta+o_y\cos\theta & a_y & a_g n_y+b_g o_y+h_g a_y \\ n_z\cos\theta+o_z\sin\theta & -n_z\sin\theta+o_z\cos\theta & a_z & a_g n_z+b_g o_z+h_g a_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{16}
$$

$$
\begin{cases} (n_z\cos\theta+o_z\sin\theta)x_{gi}x_{1cgi}+(-n_z\sin\theta+o_z\cos\theta)y_{gi}x_{1cgi}+(a_g n_z+b_g o_z+h_g a_z)x_{1cgi} \\ =(n_x\cos\theta+o_x\sin\theta)x_{gi}+(-n_x\sin\theta+o_x\cos\theta)y_{gi}+a_g n_x+b_g o_x+h_g a_x \\ (n_z\cos\theta+o_z\sin\theta)x_{gi}y_{1cgi}+(-n_z\sin\theta+o_z\cos\theta)y_{gi}y_{1cgi}+(a_g n_z+b_g o_z+h_g a_z)y_{1cgi} \\ =(n_y\cos\theta+o_y\sin\theta)x_{gi}+(-n_y\sin\theta+o_y\cos\theta)y_{gi}+a_g n_y+b_g o_y+h_g a_y \end{cases} \tag{17}
$$

Where $[n_x\ n_y\ n_z]^T$, $[o_x\ o_y\ o_z]^T$, and $[a_x\ a_y\ a_z]^T$ are computed with (6) and (7), they are the unit vectors of the X_w-, Y_w- and Z_w-axis of the world frame expressed in the camera frame. (x_{gi}, y_{gi}, z_{gi}) are the i-th point's coordinates in the target frame, (x_{1cgi}, y_{1cgi}) are the normalized imaging coordinates, θ is the orientation angle of the target in the world frame.

$$
\begin{cases} (n_z x_{gi}x_{1cgi}+o_z y_{gi}x_{1cgi}-n_x x_{gi}-o_x y_{gi})\cos\theta/h_g+(o_z x_{gi}x_{1cgi}-n_z y_{gi}x_{1cgi}-o_x x_{gi}+n_x y_{gi})\sin\theta/h_g \\ +(n_z x_{1cgi}-n_x)a_g/h_g+(o_z x_{1cgi}-o_x)b_g/h_g=(a_x-a_z x_{1cgi}) \\ (n_z x_{gi}y_{1cgi}+o_z y_{gi}y_{1cgi}-n_y x_{gi}-o_y y_{gi})\cos\theta/h_g+(o_z x_{gi}y_{1cgi}-n_z y_{gi}y_{1cgi}-o_y x_{gi}+n_y y_{gi})\sin\theta/h_g \\ +(n_z y_{1ci}-n_y)a_g/h_g+(o_z y_{1ci}-o_y)b_g/h_g=(a_y-a_z y_{1cgi}) \end{cases} \tag{18}
$$

$$
\begin{cases} (n_z x_{1cgi}-n_x)a_g+(o_z x_{1cgi}-o_x)b_g+(a_z x_{1cgi}-a_x)h_g= \\ -(n_z x_{gi}x_{1cgi}+o_z y_{gi}x_{1cgi}-n_x x_{gi}-o_x y_{gi})\cos\theta-(o_z x_{gi}x_{1cgi}-n_z y_{gi}x_{1cgi}-o_x x_{gi}+n_x y_{gi})\sin\theta \\ (n_z y_{1cgi}-n_y)a_g+(o_z y_{1cgi}-o_y)b_g+(a_z y_{1cgi}-a_y)h_g= \\ -(n_z x_{gi}y_{1cgi}+o_z y_{gi}y_{1cgi}-n_y x_{gi}-o_y y_{gi})\cos\theta-(o_z x_{gi}y_{1cgi}-n_z y_{gi}y_{1cgi}-o_y x_{gi}+n_y y_{gi})\sin\theta \end{cases} \tag{19}
$$

$$
\begin{cases} x_{g0}=a_g-a \\ y_{g0}=b_g-b \\ z_{g0}=h_g-h \end{cases} \tag{20}
$$

For the control points in the specified arrow target, their coordinates in the world frame are calculated as given (21).

$$
\begin{bmatrix} x_{gwi} \\ y_{gwi} \\ z_{gwi} \\ 1 \end{bmatrix}=\begin{bmatrix} \cos\theta & -\sin\theta & 0 & x_{g0} \\ \sin\theta & \cos\theta & 0 & y_{g0} \\ 0 & 0 & 1 & z_{g0} \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_{gi} \\ y_{gi} \\ 0 \\ 1 \end{bmatrix} \tag{21}
$$

where $(x_{gwi}, y_{gwi}, z_{gwi})$ are i-th control point's coordinates of the target in the world frame, (x_{g0}, y_{g0}, z_{g0}) are the target origin's coordinates in the world frame.

In the proposed method above, the position and orientation of the specified arrow target in the world frame can be measured with single image. The world frame is assigned with the chessboard. The chessboard can be moved away after the calibration as described in section III is finished. In addition, it should be noticed that the arrow target is on a plane parallel to the chessboard plane.

V. Experiments

The camera used in experiment had 1292×964 pixels resolution. It was fixed on the ceiling, looking down to the ground. The reference points in this experiment are the corner of the black and white chessboard with 3×3 blocks. Each block in chessboard had size of 50×50 mm. There were three arrow targets on the ground sharing almost same height with the reference points. The experiment image is shown in Fig. 2.

At first, the camera's intrinsic parameters were calibrated using Matlab calibration toolbox. The intrinsic parameters were listed as follows. k_x=1041.36089, k_y=1044.62994,

u_0=722.05012, v_0=467.05396. The lens distortion parameters were [-0.37041, 0.12123, -0.00053, -0.00715]

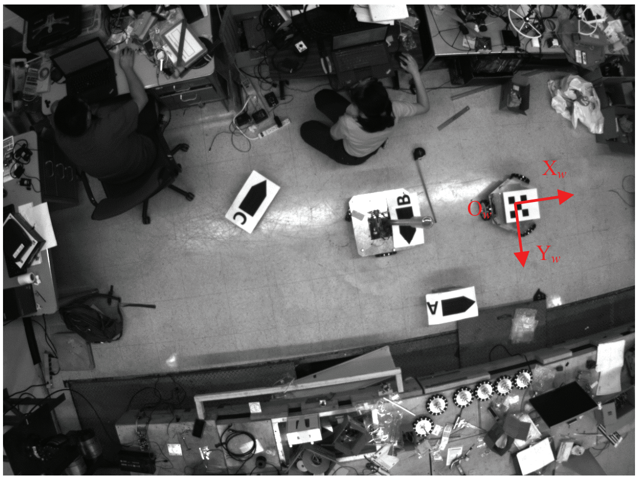

Fig. 2 Calibration scene image captured by camera

There were 16 points in the reference chessboard, which could be used for calibrating extrinsic parameters. The projection of the world frame on $O_wX_wY_w$ plane is shown in Fig. 2. The extrinsic parameters were firstly computed with the method as described in Section II. Then they were optimized under the constraint of the unit orthogonal rotation matrix. The obtained extrinsic matrix was given in (22), which was calibrated using the method discussed in section III.

$$ {}^{c}T_{w} = \begin{bmatrix} 0.9777 & 0.1784 & -0.1948 & 1084.4 \\ -0.1112 & 0.9606 & -0.1765 & -203.7 \\ 0.1781 & 0.2130 & 0.9591 & 3317.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22} $$

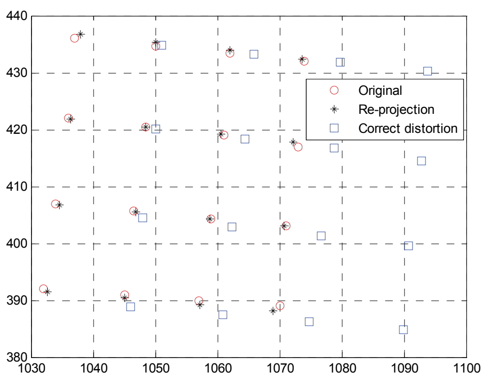

Fig. 3 The image coordinates of points in the reference Chessboard

The image coordinates of points in the reference chessboard are shown in Fig. 3. Symbol "o" denotes the original image coordinates with distortion. Symbol "□" denotes the image coordinates after distortion correction. "*" indicates the re-projected image coordinates with the extrinsic and intrinsic parameters of the camera and the planar coordinates in the reference chessboard. It can be seen that the re-projected image coordinates are coincidence with the original image coordinates. It means that the obtained extrinsic parameters in (22) are satisfactory.

Using LS method to solve (11) gave the estimated positions (a, b, h) of the world frame relative to the medium frame. The results are as follows. a=1627.1 mm, b=545.9 mm, h=3100.5 mm. Similar, the estimated positions (a_{gi}, b_{gi}, h_{gi}) and orientation angle θ_i of the i-th target frame relative to the camera medium frame were conveniently calculated using the method in section IV.B. The differences between (a, b, h) and (a_{gi}, b_{gi}, h_{gi}) are the target's position in the world frame.

In experiments with the measurement method as described in section IV.A, all the arrow tergets and reference points were considered on the same $X_wO_wY_w$ plane, sharing same height in world frame, i.e. z_{wi}=0. In experiments with the measurement method as described in section IV.B, the position variables (a_{gi}, b_{gi}, h_{gi}) and orientation angle θ_i of the i-th target frame relative to the camera medium frame were calculated. The experiment results are listed in Table I. The actual positions were manually measured. The estimated positions of vertexes on arrow targets with the measurement method as described in section IV.A were denoted as (X_{wim1}, Y_{wim1}). The estimated positions with the measurement method as described in section IV.B were denoted as (X_{wim2}, Y_{wim2}). In traditional measurement method, the camera was taken as perpendicular to the floor. The scale factor and yawing angle from image to planar coordinates was computed from the outer corners of the reference chessboard. The scale factor was 3.3611 and the yawing angle was -0.0982 rad. The estimated positions of vertexes on arrow targets with the traditional measurement method were denoted as (X_{wim3}, Y_{wim3}). It can be found that the results with the proposed method are much more accurate than ones with the traditional method.

TABLE I MEASUREMENT RESULTS AND ACTUAL POSITIONS

No.	X_{wim1} (mm)	Y_{wim1} (mm)	X_{wim2} (mm)	Y_{wim2} (mm)	X_{wim3} (mm)	Y_{wim3} (mm)	X_{wia} (mm)	Y_{wia} (mm)
1-1	-402.0	641.7	-496.3	551.0	-362.5	672.7	-436.6	584.2
1-2	-485.0	582.4	-584.4	490.1	-445.9	619.9	-526.6	526.2
1-3	-653.6	578.5	-750.3	484.8	-614.2	627.9	-692.6	526.2
1-4	-660.1	693.8	-754.0	600.7	-612.5	744.2	-692.6	642.2
1-5	-492.7	699.8	-588.1	606.1	-447.1	736.5	-526.6	642.2
2-1	-803.2	186.1	-773.9	188.5	-803.8	218.7	-838.2	173.0
2-2	-743.1	101.6	-714.1	99.6	-746.2	122.4	-780.2	83.0
2-3	-740.7	-45.1	-710.9	-66.4	-756.8	-45.3	-780.2	-83.0
2-4	-845.6	-43.4	-826.9	-68.6	-873.7	-42.7	-896.2	-83.0
2-5	-847.7	107.5	-830.1	97.3	-860.4	131.5	-896.2	83.0
3-1	-1522.6	-269.7	-1550.9	-276.6	-1745.3	-332.1	-1351.7	-402.5
3-2	-1606.6	-240.8	-1654.8	-250.5	-1853.4	-295.8	-1456.3	-379.9
3-3	-1702.4	-136.7	-1768.2	-129.3	-1960.8	-159.1	-1573.7	-262.5
3-4	-1642.7	-79.2	-1683.5	-50.0	-1865.4	-83.8	-1491.7	-180.5
3-5	-1544.4	-181.5	-1570.1	-171.3	-1755.5	-216.2	-1374.3	-297.9

Fig. 4 shows the actual and measurement positions of the three arrow targets. The arrow lines indicate the orientations of the arrow target. Symbol "□" denotes the results with the proposed method as described in section IV.B. Symbol "*" indicates the results with the method under condition z_w=0 as described in section IV.A. Symbol "+" indicates the results with the traditional planar measurement method. Symbol "o" denotes the manually measured actual positions. It's conspicuous that there exist errors between actual positions and measurement positions. These errors are mainly affected by the pose of the camera relative to the ground, the image coordinates precision of points on the reference chessboard, and the distortion of the camera's lens. The closer the arrow target to the camera center, the less errors between actual and measurement positions. Even though the position errors may be various, the orientations of the arrow targets are more accurate. The position errors and orientation errors with the proposed method are in toleration in localization task.

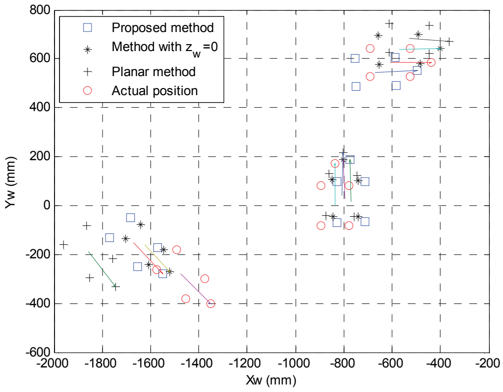

Fig. 4 Actual and measured positions of landmarks

VI. CONCLUSION

In this paper, a new localization method with high accuracy is developed based on monocular vision. With the help of reference chessboard to define the world frame, it is able to give the orientation angle and 3D position of the specified arrow target in the world frame using a single image. Besides, this method has certain toleration with the variation heights of camera. The experiments verify the effectiveness of the proposed measurement method.

In future, this method would be applied into autonomous system to estimate the positions and orientations of moving targets. In addition, how to further improve the measurement accuracy has a large space to be well investigated.

VII. ACKNOWLEDGEMENT

The authors would like to thank Associate Professor Yi and his lab for their support in this work.

REFERENCES

[1] G. N. DeSouza, A. C. Kak, Vision for mobile robot navigation: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 2, pp. 237-267, 2002.

[2] Q. Zhan, S. Huang, J. Wu, Automatic Navigation for A Mobile Robot with Monocular Vision, IEEE Conference on Robotics, Automation and Mechatronics, pp. 1005-1010, 2008.

[3] T. Goedeme, M. Nuttin, T. Tuytelaars, *et al.*, Omnidirectional vision based topological navigation, International Journal of Computer Vision, vol. 74, no. 3, pp. 219-236, 2007.

[4] H. N. Do, M. Jadaliha, J. Choi, *et al.*, Feature selection for position estimation using an omnidirectional camera, Image and Vision Computing, vol. 39, pp. 1-9, 2015.

[5] C. Mostegel, A. Wendel ; H. Bischof, Active monocular localization: Towards autonomous monocular exploration for multirotor MAVs, IEEE International Conference on Robotics & Automation, pp. 3848-3855, 2014.

[6] A. Wendel, A. Irschara ; H. Bischof, Natural landmark-based monocular localization for MAVs, IEEE International Conference on Robotics & Automation , pp. 5792 – 5799, 2011.

[7] Y. J. Yin, D. Xu, Z. T. Zhang, *et al.*, Plane measurement based on monocular vision, Journal of Electric Measurement and Instrument, vol. 27, no. 4, pp. 347-352, 2013. (in Chinese)

[8] D. Xu; L. Han; M. Tan, *et al.*, Ceiling-Based Visual Positioning for an Indoor Mobile Robot With Monocular Vision, IEEE Transactions on Industrial Electronics, vol. 56, no. 5, pp. 1617 – 1628, 2009.

[9] X. Huang, F, Gao, G. Xu, *et al.*, Depth information extraction of on-board monocular vision based on a single vertical target image, Journal of Beijing University of Aero-nautics and Astronautics, vol. 41, no. 4, pp. 649-655, 2015. (in Chinese)

[10] R. C. A. Nascimento, B. M. F. Silva, L. M. G. Gonçalves, Real-time localization of mobile robots in indoor environments using a ceiling camera structure, IEEE Latin American Robotics Symposium, pp. 61-66, 2013.

[11] W. Zhu, P. Wang, F. Li, *et al.*, Real-time 3D Model-based Tracking of Work-piece with Monocular Camera, IEEE/SICE International Symposium on System Integration (SII), pp. 777- 782, 2015.

[12] Q. X. Yu, W. X. Yan, Z. Fu, Y.Z. Zhao, Service robot localization based on global vision and stereo vision, Journal of Donghua University (English Edition), vol. 29, no. 3, pp. 197-202,2012.