A New 2-Correspondences Monocular Vision Navigation Method Under Planar Motion

Yingjian Yu$^{1,2}$, Xiangyi Sun$^{1,2}$ and Banglei Guan$^{1,2*}$

$^1$College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, Hunan, 410073, China

$^2$Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, National University of Defense Technology, Changsha, Hunan, 410073, China

$^*$Corresponding author’s e-mail: banglei.guan@hotmail.com

Abstract. Vision navigation has non-contact, high-precision and low-cost characteristics compared with other navigation method. This paper proposes a vision navigation method for planar moving platform in known scene based on two 2D-3D correspondences. Taking account of the arbitrary direction of camera and the relative pose between the camera and the platform, we propose a linear solver with unique solution, whose minimal number of requirement correspondences is 2 and can cope with arbitrary number of correspondences bigger than 2 directly. Our method is evaluated on both synthetic data and real-world data. The experimental results demonstrate that our method is high-accuracy and fast for vision navigation of planar moving platform.

1. Introduction

Vision navigation is the technology to determine the position, velocity, orientations or other navigation information of a moving platform by using the visible-light digital camera, infrared camera or imaging radar on it to image the scene, thus offering the navigation information to the moving platform[1]. Vision navigation method is famous for its non-contact, high-precision and low-cost feature, which has been widely concerned in recent years[2-5]. The most frequently used vision sensor in vision navigation is visible-light digital camera (camera for short)[6]. Compared with the laser and radar navigations, the camera-based vision navigation (vision navigation for short) only needs cameras, and it does not need to actively emit laser and electromagnetic waves to the target and receive the returned information. Compared with Global Positioning System (GPS), visual navigation is not limited by the coverage of satellite signal, and has very strong anti-jamming performance, while radio and GPS signals are easy to be interfered and blocked. Visual navigation can be carried out indoors or even underground, and there is no need to establish satellite navigation system and develop signal receiving device, so the cost is low. With the development of visual navigation technology, the need for automated task based on vision, such as allocation and transportation in warehouses and docks, is urgent, where the navigation scene such as warehouses and docks are specific, known and generally in a plane. In this paper, we focus on the visual navigation method on the plane in known scene.

The key of vision navigation is solving Perspective-n-Point (PnP) problem, whose goal is to determine the absolute pose (rotation and translation) of a calibrated camera in a world reference frame, given known three-dimensional (3D) world points and corresponding two-dimensional (2D) image observations, i.e. 2D-3D correspondences. If camera intrinsic parameters matrix is unknown, the
absolute pose estimation is the same as camera resectioning in the photogrammetric. In general, the camera pose has 6 degrees-of-freedom (DOF), namely the rotation (roll, pitch, and yaw) and 3D translation of the camera with respect to the world. Generally, three 2D-3D correspondences are enough to solve the PnP problem, where up to four real, geometrically feasible solutions are yielded. For low noise levels, a fourth point can be used for disambiguation[7]. Other methods with fixed point correspondences are P4P[8] and P5P[9]. Those methods can not directly used for arbitrary number of points correspondences. However, many other methods can be easily used for arbitrary number of correspondences. Basically, they can be categorized into iterative, non-iterative or polynomial, non-polynomial solvers, including the MLpNP (maximum likelihood PnP solution)[10], LHM (one of the best iterative solutions of PnP)[11], the EPnP+GN (efficient O(n) noniterative PnP solution together with a few Gauss-Newton steps)[12], the RPnP (robust O(n) PnP method)[13], the DLS (direct least square solution)[14], the PPNpN (PnP problem with anisotropic orthogonal procrustes analysis)[15], the ASPnP (accurate and scalable PnP)[16], the SDP (globally optimal O(n) Solution to the PnP problem for general camera models)[17], the OPNpN (optimal solution to the PnP problem)[18], and the EPPnP (Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection)[19]. When there are outliers existing in correspondences, the random sample consensus (RANSAC)[20] scheme is used to eliminate outliers.

However, when the camera’s motion is a planar motion, the DOF reduces to 3. Therefore, the vision navigation for planar-motion object is to measure the yaw and planar coordinate (X, Y) in world frame. The minimal number of correspondences required is two. Choi et al. [21] proposed an effective minimal solution for absolute pose estimation of platform under planar motion based specific simplification and special assumption. In [21], the relative pose (position and orientations) between camera and platform is ignored, however, without which the absolute pose of the platform can’t be recovered. Therefore the method in [21] can’t solve the navigation problem directly, which only estimate the pose of camera instead. Additionally, the X-Z plane of the camera coordinate system has to be parallel with the moving plane of the platform. In practice, however, the alignment error has an adverse effect to the final result and sometimes we need camera to look sideways or in other direction. In this paper, we take account of the arbitrary direction of camera and the relative pose between the camera and the platform, and propose a linear solver with unique solution, whose minimal number of requirement correspondences is 2 and can cope with arbitrary number of correspondences bigger than 2 directly.

The remainder of the paper is organized as follows. In Section 2, we propose a 2-Correspondences monocular vision navigation method under planar motion. In Section 3, we evaluate the performance of proposed methods using both synthetic and real-world dataset. Finally, concluding remarks are given in Section 4.

2. Method

2.1 Basics and notations

In this paper, the pinhole camera model is used to model the perspective projection mapping from a 3D world point to a 2D image point. Figure 1 shows a basic geometrical constraint of the proposed approach. A platform with a camera on it is moving on the plane W-XY. The world coordinate system is W-XYZ, and \( P_1(X_1, Y_1, Z_1) \), \( P_2(X_2, Y_2, Z_2) \) are 3D world point in W-XYZ. The camera coordinate system is C-Xc,Yc,Zc while the platform coordinate system is B-Xb,Yb,Zb. The rotation matrix from B-Xb,Yb,Zb to C-Xc,Yc,Zc is \( R_{cb} \) and the camera’s optical centre is C whose coordinate in B-Xb,Yb,Zb is \( t_{bc} = (x_c, y_c, z_c)^T \). The image pixel coordinate system is I-xy, while the image coordinate system is O-x'y'. where the point O is the principal point and the point I is the left upper corner of image. The image observations \( p_1(x_1, y_1) \) and \( p_2(x_2, y_2) \) described in I-xy is corresponding to \( P_1(X_1, Y_1, Z_1) \) and \( P_2(X_2, Y_2, Z_2) \), respectively.
Figure 1. Geometric model of monocular vision navigation for planar moving objects

The position of the platform in W-XYZ is \( t_{wb} = (X, Y, H) \) and the platform’s yaw is \( \theta \). Since the motion is on the plane \( W-XY \), \( H \) is a constant and the remaining angle (roll and pitch) is zero. The rotation matrix from \( W-XYZ \) to \( B-B \) is

\[
R_{wb} = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}.
\]  

(1)

The relation between 2D image point \( p_i = (x_i, y_i)^T \) and 3D world point \( P_i = (X_i, Y_i, Z_i)^T \) can be expressed as:

\[
\begin{bmatrix}
p_i \\
1
\end{bmatrix} \approx KR_{cb}(R_{wb}(P_i - t_{wb}) - t_{bc}),
\]  

(2)

where \( i \) is the sequence number of the 2D-3D correspondence and \( K \) is the intrinsic parameters matrix of the camera.

2.2 The 2-Correspondences method

The Eq. (2) describes the relation between 2D image point and 3D world point and it take into account the transformation between the camera and the platform. The goal of our method is to find the yaw \( \theta \) (in \( R_{wb} \)) and planar coordinate \( (X, Y) \) in world frame, since the relation between camera and the platform i.e. \( R_{cb} \) and \( t_{bc} \) can be determined in advance. The \( \theta \) and \( (X, Y) \) will be coupled together if we directly to solve Eq. (2), since \( R_{wb}(P_i - t_{wb}) \) exits in Eq. (2). Therefore, we formulate the relation in Eq. (2) in another from:

\[
\begin{bmatrix}
p_i \\
1
\end{bmatrix} \approx K(R_{cb}(R_{wb}P_i + t_{wb}) + t_{cb}),
\]  

(3)

where \( t_{cb} = -R_{cb}t_{bc} \), and \( t_{wb} = [X_b, Y_b, Z_b] \) is the coordinate of world coordinate system’s origin, which is expressed in platform coordinate system and \( Z_b = -H \). Then the Eq. (3) can be expressed as:
where \( \mathbf{p}_i = K^{-1} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \) is the \( i \)th normalized image point and \( r_{jk} (j, k = 1, 2, 3) \) is the \( j \)th row and \( k \)th column element of \( \mathbf{R} \). Finally, Eq. (4) is converted to a linear system

\[
\mathbf{AV} = \mathbf{L},
\]

where

\[
\mathbf{A} = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[
\mathbf{V} = \begin{bmatrix}
\cos \theta \\
\sin \theta \\
X_B \\
Y_B
\end{bmatrix},
\]

\[
\mathbf{L} = \begin{bmatrix}
t_1 + Z_{r_{13}} - r_{13} H - \tilde{x}_1 (t_3 + Z_{r_{33}} - r_{33} H) \\
t_2 + Z_{r_{23}} - r_{23} H - \tilde{y}_1 (t_3 + Z_{r_{33}} - r_{33} H)
\end{bmatrix}_{i=1,2,\ldots,n}
\]

and \( t_m \) is the \( m \)th \( (m=1,2,3) \) element of \( \mathbf{t}_{CB} \).

Two 2D-3D correspondences are the minimal configuration to get the solution. When there are more correspondences, the solution can be acquired by using the least square minimization. After solving Eq. (5) using singular value decomposition (SVD) method, we acquire the value of variables in \( \mathbf{V} \). However, the acquired value \( (v_1, v_2) \) of \( \cos \theta \) and \( \sin \theta \) generally do not satisfy the theorem:

\[
\sin^2 \theta + \cos^2 \theta = 1.
\]

To adjust this error, we determine the value of \( \theta \) by follow equation:

\[
\theta = \frac{\arctan \left( \frac{\sqrt{1-v_1^2}}{v_1} \right) + \arctan \left( \frac{v_2}{\sqrt{1-v_2^2}} \right)}{2}.
\]

And the intermediate result the final result of platform’s position are

\[
\mathbf{t}_{BW} = \begin{bmatrix}
X_B \\
Y_B \\
-H
\end{bmatrix},
\]

\[
\mathbf{t}_{WB} = -\mathbf{R}_{BW} \mathbf{t}_{BW},
\]

where the coordinate \( (X,Y) \) of the platform in plane \( W-XY \) is the first two elements in \( \mathbf{t}_{WB} \).

3. Results and Discussion

In this section, synthetic as well as real data is used to estimate our method LP2P (linear 2-correspondences vision navigation method). We compare our algorithm to ten state-of-the-art algorithms in both accuracy and time consumption. All experiments have been conducted in MATLAB R2016a on a Laptop with Intel Core i5-8250U CPU @1.6Ghz.

3.1 Synthetic experiments

In this section, we use synthetic data to estimate the performance of LP2P and compare it with state-of-the-art algorithms including the MLPnP (maximum likelihood PnP solution)[10], LHM (one of the best
iterative solutions of PnP\cite{11}, the EPnP+GN (efficient O(n) noniterative PnP solution together with a few GaussNewton steps)\cite{12}, the RPnP (robust O(n) PnP method)\cite{13}, the DLS (direct least square solution)\cite{14}, the PPnP (PnP problem with anisotropic orthogonal procrustes analysis)\cite{15}, the ASPnP (accurate and scalable PnP)\cite{16}, the SDP(globally optimal O(n) Solution to the PnP problem for general camera models)\cite{17}, the OPnP (optimal solution to the PnP problem)\cite{18}, and the EPPnP (Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection)\cite{19}.

In order to facilitate the experimental comparison and verification, we unified the experimental configuration to generate data for simulation. The random relative pose (rotation and translation) between the virtual camera and the virtual planar-moving platform is generate. Then random position and yaw with respect to the world coordinate system is generate. The virtual camera with a focal length of 800 pixels and image resolution of $640 \times 800$ is used to generate synthetic data, whose principle point is exactly in the center of the image. The range of coordinates in camera space is $[-2,2] \times [-2,2] \times [1,8]$. According to different configurations, several sets of different target 3D reference points and corresponding 2D image points of the dataset are generated.

In order to evaluate the accuracy of the position and the yaw, we use the mean, median and mean reprojection error to quantitatively evaluate LP2P and compare other with state-of-the-art algorithms. These evaluation criterions are defined as follow:

\[
\bar{\theta}_{\text{yaw}} = \text{mean}(\|\theta - \theta_{\text{true}}\|) = \frac{1}{n} \sum_{k=1}^{n} \|\theta^k - \theta_{\text{true}}\|,
\]

\[
\check{\theta}_{\text{yaw}} = \text{median}(\|\theta - \theta_{\text{true}}\|),
\]

where $\theta$ and $\theta_{\text{true}}$ are vectors of estimated yaw values and ground truth of all iterations, respectively, $n$ is the total number of iterations;

\[
\bar{\theta}_{\text{position}} = \text{mean}(\|\mathbf{L} - \mathbf{L}_{\text{true}}\|) = \frac{1}{n} \sum_{k=1}^{n} \|\mathbf{L}^k - \mathbf{L}_{\text{true}}\|,
\]

\[
\check{\theta}_{\text{position}} = \text{median}(\|\mathbf{L} - \mathbf{L}_{\text{true}}\|),
\]

where $\mathbf{L}$ and $\mathbf{L}_{\text{true}}$ are vectors of estimated positions and ground truth of all iterations, respectively, $n$ is the total number of iterations;

\[
\bar{\varepsilon}_{\text{repro}} = \text{mean}(\|R \ t \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix} - \mathbf{P}_c\|),
\]

\[
\check{\varepsilon}_{\text{repro}} = \text{median}(\|R \ t \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix} - \mathbf{P}_c\|),
\]

where $R$ and $t$ is the rotation matrix and the translation vector from world coordinates to camera coordinates of 3D points, $\mathbf{P}_c$ is the coordinate matrix with its every column of 3D point’s coordinate in the camera space, and $\mathbf{P}$ is the coordinate matrix of 3D points in world coordinate system. Note that both $R$ and $t$ are different in different iteration. The criterion for determining time consumption is the mean time consumption of every iteration and we use $\bar{\tau}$ to represent it.
Figure 2. The LP2P’s mean errors and median errors of the estimated yaw and position comparing with other methods.

In this experiment, the accuracy of proposed method (LP2P) is evaluated in different noise level. The generated image point coordinates are added random Gaussian noise whose mean is zero and maximal standard deviation is from 1 to 10 pixels with the step of 1. Two thousand repeated experiments are implemented in every noise level. Figure 2 shows the experiment results in the yaw estimation and position estimated, separately. As shown in Figure 2 (a) and (b), the mean and median errors of yaw of LP2P method are small even when the noise level is up to 10 pixels and the maximal median errors of LP2P is smaller than 0.015, even though there are other methods outperform LP2P method. As shown in Figure 2 (c) and (d), the LP2P method outperforms all methods in mean position errors in all noise level, and it outperforms almost all method in median position errors in all noise level except for MLPnP method.
Figure 3. The mean (a) and median (b) reprojection errors of LP2P comparing with other methods.

Figure 4. The mean (a) and median (b) time consumption of LP2P comparing with other methods.

For comprehensive accuracy evaluation, we evaluate all methods in the reprojection errors defined in Eq. (11). As shown in Figure 3, LP2P has comparable accuracy in criterion of mean reprojection errors with SDP, ASPnP and OPnP, which outperforms the rest of methods. Further, LP2P together with MLPnP is apparently most accuracy than all other methods in the criterion of median reprojection errors. In summary, LP2P is of greater high accuracy and relative stability than all compared methods. Figure 4 shows the time consumption of all methods using six correspondences to estimate pose. LP2P method has the least time consumption than the rest method. Note that: for clearer comparison in Figure 2, 3 and 4, we reduce the maximum of the scope of the y-axis, which may lead to the results of higher errors can’t be displayed in the Figure.

The time consumption of different methods with the scale of correspondences increasing is evaluated. We vary the number of pairs of correspondence from 10 to 100 with the step of 10, repeat 1000 times in every scale and plot the mean time consumption in Figure 5. Since the number of control point in an image in practice is usually less than 100, we set the maximum of test point correspondence as 100.
experiment results in Figure 5 demonstrate that the proposed LP2P mean is less time-consuming. And LP2P is apparently faster than all other methods in case that the number of 2D-3D correspondences is less than 20.

![Figure 5.](image)

**Figure 5.** The mean (a) and median (b) time consumption of LP2P comparing with other methods.

### 3.2 Experiments with real data

To further validate the effectiveness of the proposed method, we record a sequence of images with the resolution of $2448 \times 2048$ captured by the camera fixed on a moving platform in plane. The number of images in the trajectory is $M = 100$. In every image, the ground truth pose of platform is measured by Leica total station whose ranging accuracy is 0.1 mm and angle measurement accuracy is 0.5 arcsecond. The coordinates of 3D points are measured by total station.

![Image](image)

**Figure 6.** The moving platform with camera is in (a) and the image captured by camera is in (b).

The moving platform in this experiment is shown in Figure 6 (a) and the image captured by camera is in Figure 6 (b). A series of markers in experiment scene are set for tracking whose coordinates are determined by total station. We extract the image coordinates of these markers in image by using the automatic extraction method of diagonal markers in [22]. By matching 2D image coordinates of markers
with corresponding 3D world coordinates, we get the 2D-3D correspondences. Since the relation between camera and platform has been calibrated in advance, the pose (yaw $\theta$ and planar coordinates $(X, Y)$) of platform can be estimated. The processing is conducted offline. To evaluate the accuracy, we compare the estimated pose with the ground truth. As shown in Table 1, our method has the best performance in errors of position.

| Method         | Yaw (deg.) $\varepsilon_{yaw}$ | Yaw (deg.) $\tilde{\varepsilon}_{yaw}$ | Position (m) $\varepsilon_{position}$ | Position (m) $\tilde{\varepsilon}_{position}$ |
|----------------|---------------------------------|------------------------------------------|----------------------------------------|-----------------------------------------------|
| LP2P           | 0.0163                          | 0.0154                                   | 0.0210                                 | 0.0158                                        |
| MLPnP[10]      | 10.373                          | 10.37                                    | 14.96                                  | 12.927                                        |
| LHM[11]        | 1.782                           | 1.782                                    | 0.2563                                 | 0.2218                                        |
| EPnP+GN[12]    | 0.0284                          | 0.0284                                   | 0.7668                                 | 0.0258                                        |
| RPnP[13]       | 0.0431                          | 0.0431                                   | 0.0287                                 | 0.0267                                        |
| DLS[14]        | 0.0025                          | 0.0025                                   | 0.0221                                 | 0.0187                                        |
| PPnP[15]       | 0.0152                          | 0.0152                                   | 0.0225                                 | 0.0194                                        |
| ASPnP[16]      | 0.0024                          | 0.0053                                   | 0.0221                                 | 0.0188                                        |
| OPnP[18]       | 0.0081                          | 0.0023                                   | 0.0200                                 | 0.0187                                        |
| EPPnP[19]      | 109.20                          | 109.2                                    | 9.020                                  | 8.937                                         |

The time consumption is shown in Table 2. Our LP2P has the least time consumption, which demonstrates that the LP2P is fast.

| Method         | $\tau$ (ms) |
|----------------|-------------|
| LP2P           | 0.56        |
| MLPnP          | 3.43        |
| LHM            | 4.56        |
| EPnP+GN        | 2.22        |
| RPnP           | 1.60        |
| DLS            | 10.43       |
| PPnP           | 8.41        |
| ASPnP          | 3.40        |
| OPnP           | 13.53       |
| EPPnP          | 0.91        |

Since SDP is failed in real experiment, we can’t report its result.

4. Conclusions
The proposed LP2P method for a vision navigation method for planar moving object in known scene based on two 2D-3D correspondences is effective and fast. We take account of the arbitrary direction of camera and the relative pose between the camera and the platform, and propose a linear solver with unique solution, whose minimal number of requirement correspondences is 2 and can cope with arbitrary number of correspondences bigger than 2 directly. Our method is evaluated on both synthetic data and real-world data. The experimental results demonstrate that our method is both high-accuracy and fast for vision navigation of planar moving platform. The proposed method can speed up the process of vision navigation of planar moving platform and has a high accuracy. By considering the arbitrary direction of camera and the space relation between the camera and the platform, the LP2P makes the application of vision navigation go forward further.

The proposed method is restricted in planar-motion which is common in real work scene. The future research includes developing the applications for automatic driving by considering the plane error.

Acknowledgments
The research was supported by National Natural Science Foundation of China (No.11902349).

References
[1] Yu, Q., and Shang, Y. (2009) Videometrics: Principles and Researches. Science Press.
[2] Guan, B., Vasseur, P., Demonceaux, C., and Fraundorfer, F. (2018) Visual Odometry Using a
Homography Formulation with Decoupled Rotation and Translation Estimation Using Minimal Solutions. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). 2320-2327.

[3] Guan, B., Zhao, J., Li, Z., Sun, F., and Fraundorfer, F. (2020) Minimal Solutions for Relative Pose With a Single Affine Correspondence. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 1926-1935.

[4] Guan, B., Zhao, J., Li, Z., Sun, F., and Fraundorfer, F. (2021) Relative Pose Estimation With a Single Affine Correspondence. IEEE Transactions on Cybernetics1-12.

[5] Tian, M., Guan, B.L., Xing, Z.B., and Fraundorfer, F. (2020) Efficient Ego-Motion Estimation for Multi-Camera Systems With Decoupled Rotation and Translation. Ieee Access, 8: 153804-153814.

[6] Gao, X., Zhang, T., Liu, Y., and Yan, Q. (2017) 14 Lectures on Visual SLAM: From Theory to Practice. Publishing House of Electronics Industry, Beijing.

[7] Li, S., and Xu, C. (2011) A Stable Direct Solution of Perspective-three-point Problem. International Journal of Pattern Recognition and Artificial Intelligence, 25: 627-642.

[8] Liu, M.L., and Wong, K.H. (1999) Pose Estimation Using Four Corresponding Points. Pattern Recognition Letters

[9] Zhang, Z., Sun, C., and Wang, P. (2012) Two-step pose estimation method based on five reference points. Chinese Optics Letter, 010: P.52-56.

[10] Urban, S., Leitloff, J., and Hinz, S. (2016) Mlpnp - a Real-Time Maximum Likelihood Solution to the Perspective-N-Point Problem. Isprs Ann Photo Rem, 3: 131-138.

[11] Lu, C.P., Hager, G.D., and Mjolsness, E. (2000) Fast and globally convergent pose estimation from video images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22: 610-622.

[12] Lepetit, V., Moreno-Noguer, F., and Fua, P. (2008) EPnP: An Accurate O(n) Solution to the PnP Problem. International Journal of Computer Vision, 81: 155-166.

[13] Li, S.Q., Xu, C., and Xie, M. (2012) A Robust O(n) Solution to the Perspective-n-Point Problem. Ieee Transactions on Pattern Analysis and Machine Intelligence, 34: 1444-1450.

[14] Hesch, J.A., and Roumeliotis, S.I. (2011) A Direct Least-Squares (DLS) method for PnP. In: 2011 International Conference on Computer Vision. 383-390.

[15] Garro, V., Crosilla, F., Fusiello, A., and Soc, I.C. (2012) Solving the PnP Problem with Anisotropic Orthogonal Procrustes Analysis.

[16] Zheng, Y., Sugimoto, S., and Okutomi, M. (2013) ASPnP: An Accurate and Scalable Solution to the Perspective-n-Point Problem. Ieee Transactions on Information and Systems, E96D: 1525-1535.

[17] Schweighofer, G., and Pinz, A. (2008) Globally Optimal O(n) Solution to the PnP Problem for General Camera Models. In: Proceedings of the British Machine Vision Conference 2008, Leeds, September 2008.

[18] Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., and Okutomi, M. (2013) Revisiting the PnP Problem: A Fast, General and Optimal Solution. In: 2013 IEEE International Conference on Computer Vision. 2344-2351.

[19] Ferraz, L., Binefa, X., Moreno-Noguer, F., and Ieee. (2014) Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection. In: (Eds.), 2014 Ieee Conference on Computer Vision and Pattern Recognition. Ieee, New York. 501-508.

[20] Mach, C. (1981) Random Sample Consensus: a paradigm for model fitting with application to image analysis and automated cartography. In.

[21] Choi, S.-I., and Park, S.-Y. (2015) A new 2-point absolute pose estimation algorithm under planar motion. Advanced Robotics, 29: 1005-1013.

[22] Wang, G., Shi, Z., Shang, Y., and Yu, Q. (2018) Automatic Extraction of Diagonal Markers Based on Template Matching and Peaks of Gradient Histogram. Acta Optica Sinica, 38: 156-163.