Geodetic Coordinate Calculation Based on Monocular Vision on UAV Platform

Zhi Li*, Tao Yang*, Guangpo Li*, Jing Li‡ and Yanning Zhang*
*School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710129, PR China
‡School of Telecommunications Engineering, XiDian University, Xi'an 710071, PR China
zLeewack@163.com, tyang@nwpu.edu.cn, liguangponwpu@gmail.com, jinglixd@mail.xidian.edu.cn, ynzhangnwpu@gmail.com

***Abstract*—Vision measurement technology research is active in recent years, which has broad application prospect in the fields of military and civilian industrial production. On the basis of comparing several existing measurement techniques, our paper mainly studied a method of geodetic coordinate calculation based on monocular vision on UAV platform, which is extension and expansion of the traditional monocular vision measurement method. Our system only needs a monocular camera carried on UAV platform and combines the latest ORB-SLAM algorithm, we can virtually extend the monocular camera into the binocular camera and even the multi-view camera system. What about the virtuality? When we perform the ORB-SLAM, our system can accurately estimate the camera pose information of each KeyFrame, it is conceivable that we can view each KeyFrame as a separate camera view, and then we select the correlative frames of the target to calculate the geodetic coordinate optimally using multiple views projective reconstruction method. All outcomes of indoor and outdoor experiments show that our system has good measuring precision, the details shown in the following sections.**

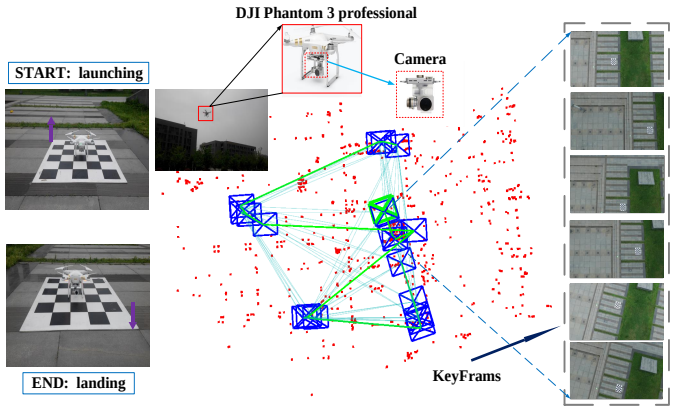

Fig. 1. Data acquisition and data processing based on the ORB-SLAM of our system. **Left:** the UAV launching and landing processes. **Middle:** the DJI Phantom 3 Professional and its built-in camera are selected as our UAV platform and monocular camera, the red points and blue boxes represent the map points and the camera keyFrames poses respectively estimated by ORB-SLAM. **Right:** the corresponding KeyFrames.

I. Introduction

With the fast development of the mobile machinery, including mobile robots and unmanned aerial vehicle (UAV), etc., measurement technology has become a hot research field in recent years. Nowadays, the most common measurement technologies usually include two categories: sensor-based measurement technology and vision-based measurement technology. For sensor-based method [1], [2], many different sensors, such as satellite, laser, ultrasonic, etc., have been used to aid the varieties of mobile devices to measure and perceive their surroundings, like the self-driving cars based on the radar and laser assistance. Vision-based method [3], [4], [5], [6], [7], [8], [9], [10], [11], which mainly research the mapping relationships of the information from 2D to 3D cartesian space according to the vision information captured from the cameras, and then assist to calculate the 3D geodetic coordinate of the target using camera imaging principle and various computer vision algorithms.

In category one, some common methods, try to get the accurate 3D geodetic coordinate depending on some special sensor absolutely, like satellite, laser, ultrasonic, which have high measurement accuracy actually, but unfortunately, these sensors would be easily affected in the environment without GPS or are susceptible to signals interference, such as indoor, underground and other shield places. However, our monocular vision-based method naturally keeps our system from affecting by the above cases.

In category two, one of the most effective method is binocular stereo vision measurement, which works pretty well for fixed scenes and has a lot of applications [6], [11]. While some drawbacks such as smaller field of measurement view, hard stereo matching limit its application in the mobile measurement platform. As for the traditional monocular camera method in[12], the author proposed a positioning method with the intersection of line and plane is given using the basic theory of imaging, the biggest limitation of which is lacking the depth scale.

The above problem can be easily addressed in our system. Primarily, our system has a simple structure, we select the DJI Phantom 3 Professional as our mobile platform and the built-in camera as our monocular camera sensor see Figure 1. Further, it is clear that monocular camera calibration process is much easier than the binocular stereo camera calibration, and the most important, our mobile platform significantly give camera a more larger field of view, which can effectively solve the limitations of the camera view that stereo camera exists.

Recently, there is a rapidly literature on monocular vision SLAM, Simultaneous Localization and Mapping (SLAM) [13], [14] is a hot field of research in computer vision due

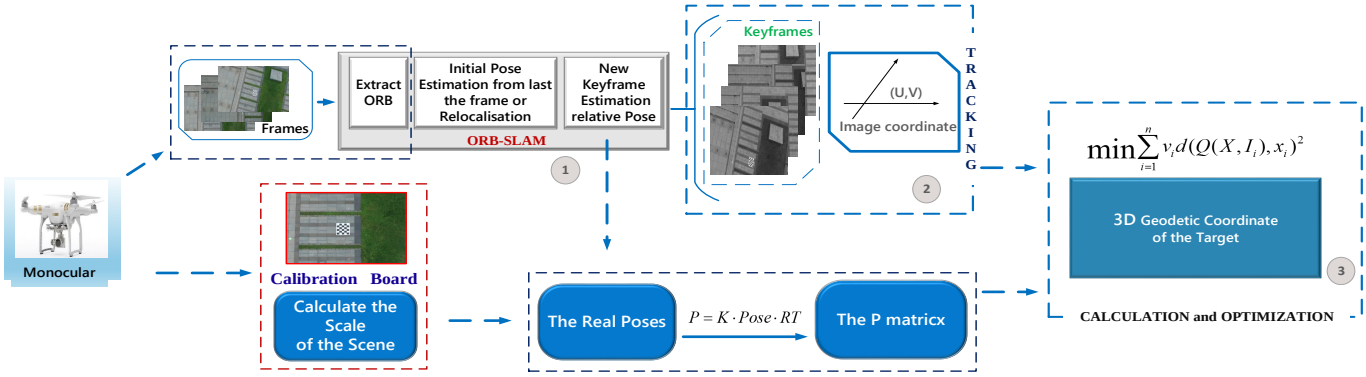

Fig. 2. The framework of our system, showing the main three modules: (1) the camera pose estimation based on ORB-SLAM and the scale calculation based on calibration board, (2) tracking the target in all the KeyFrames and getting corresponding image coordinates, (3) geodetic coordinate calculation of the target and further optimization.

to its significant applications to next-generation technologies in robotics and augmented reality [15], [16], [17], [18], [19], [20]. In our paper, we experiment with a lean version of ORB-SLAM based on the foundations laid by ORB-SLAM[14], [21], [22], [23], [24], [25], which demonstrates its great capability in localization, we attempt to use it to obtain the relative pose of each KeyFrame.

In our paper, we focus on the monocular camera mounted on the mobile platform, and the framework of our system see figure 2. Our system makes three key contributions:

- We built a original measurement system, which can virtually extend the monocular camera into multiple camera views successfully using ORB-SLAM, it is the basic and crucial step for our method to make further calculation in the following steps.
- We make the image coordinate acquisition of the target as a tracking problem. We added an extra and independent tracking module in our lean ORB-SLAM algorithm, which purpose is to track the target in the original video sequences, and then we record the image coordinates of the target whenever a new KeyFrame inserts.
- As we all know, monocular vision inherently exists the weakness of the scale of uncertainty. However, our method effectively solve the problem by using the calibration board, which help to calculate the scale in advance, then we can easily change the relative poses to real poses.

II. Calculate 3D coordinate

In this section, we mainly introduce the framework of every module in our system and the algorithm process in details.

As we can see in the figure 2, there are three main modules in our system: **orb-slam**, **tracking**, **calculation and optimization**. The first module, we deal with the original video with our lean version of ORB-SLAM, we can firstly obtain the relative poses of all the KeyFrames, and then we calculate the scale of the scene which can help to convert the relative poses into the real poses. The second module, we track the target in the original video sequences, which can easily get the image coordinates of each KeyFrame. Based on the above two modules, we have obtained the true poses of all the KeyFrames and the corresponding image coordinates of the target. And in the last module, we calculate the 3D geodetic coordinate of the target using multiple views projective reconstruction method, and make further optimization by the most popular method bundle adjustment(BA) [20], [26], [27], which has good effects.

A. Estimate Camera Pose with ORB-SLAM

ORB-SLAM[14] is a very recent paper in SLAM, which is one of the most successful feature-based SLAM methods to date. On the basic of it, we built an abridged version, and then accurately estimate the relative camera poses of all KeyFrames. As we all know, the scale uncertainty is a inherent fault in monocular camera, but fortunately, in the ORB-SLAM, among all of the scales are unified, which inspired us to use the calibration board to calculate the scale of the scene. We firstly use the camera calibration program developed by OPENCV to calibrate the parameters of five KeyFrames $\{f_1, f_2, f_3, f_4, f_5\}$ selected uniformly in advance, in which the f_1 is the reference KeyFrame in the whole process, then we can calculate five transformation matrix $\{P_1, P_2, P_3, P_4, P_5\}$. Secondly, we calculate the true poses $\{Pose_2, Pose_3, Pose_4, Pose_5\}$ by:

$$Pose_i = P_i P_1^{-1} \tag{1}$$

Where $i = 2, 3, 4, 5$. Finally, we can easily get the scale of the scene by comparing the relative poses and the true poses.

B. Tracking the Target

We make the image coordinate calculation of the target as a tracking problem, the purpose of which is to obtain the image coordinates of the target from the KeyFrame sequences. In our method, we adopt one of the most successfully tracking algorithms STC to deal with the problem, which is embedded in our lean ORB-SLAM algorithm.

In our experiment, we test it in different scenes, it has both good performance in tracking the target we selected in advance, and the results are consistent with the test requirement.

C. Calculation and Optimization

As we talk in section 1, our system virtually extend the monocular camera into multiple view cameras. In order to get the initial value of the 3D coordinate of the target, we primarily use the projective invariant of binocular method based on the minimum weight reprojection error, and the method to find the exact solution to meet the minimum polar geometric constrain and reprojection error. Since the whole process only involves the projection of the space point and the distance between 2D image points, which is invariant, that is to say, the method has nothing to do with the special projective space.

There are observation points x and x' on the images of C and C', set the $\hat{x}$ and $\hat{x}'$ as the points which exactly meet corresponding geometric constrains near the observation point, on the basic of the binocular positioning method of minimum reprojection error, we calculate $\hat{x}$ and $\hat{x}'$ by minimizing the follow function:

$$C(\hat{x}, \hat{x}') = d(x, \hat{x})^2 + d(x', \hat{x}')^2 \tag{2}$$

with $\hat{x}'^T F \hat{x} = 0$.

Where F is a fundamental matrix, which we calculate with a stereo pair of frames, $d(x, \hat{x})^2$ and $d(x', \hat{x}')^2$ are the reprojection errors computed using the corresponding projection matrices.

The process of solving the above equation is divided into two steps. Firstly, we use the DLT(Direct Linear Transform) to get the initial value of $\hat{x}$ and $\hat{x'}$, and then we optimize the $\hat{x}$ and $\hat{x'}$ using Levenberg-Marquardt's iterative non-linear optimization. Set $x \cong PX$ and $x' \cong P'X$, with the homogeneous relations $x \times PX = 0$, $x' \times P'X = 0$, there are:

$$\begin{aligned} x_1(P^{3T}X) - (P^{1T}X) &= 0 \\ y_1(P^{3T}X) - (P^{2T}X) &= 0 \\ x_1(P^{2T}X) - y_1(P^{1T}X) &= 0 \\ x_2(P'^{3T}X) - (P'^{1T}X) &= 0 \\ y_2(P'^{3T}X) - (P'^{2T}X) &= 0 \\ x_2(P'^{2T}X) - y_2(P'^{1T}X) &= 0 \end{aligned} \tag{3}$$

Where P^{iT} is the i-th row of the matrix P, P^{jT} is the j-th col of the matrix P', homogenous coordinates $x = (x_1, y_1, 1)$, $x' = (x_2, y_2, 1)$. The above is linear equations with respect to X, which can be written as $AX = 0$, although each point corresponding to the three equations, of which only two are linearly independent, and therefore each point just gives two equations correspondingly with respect to X, thus A becomes:

$$A = \begin{bmatrix} x_1 P^{3T} - P^{1T} \\ y_1 P^{3T} - P^{2T} \\ x_2 P'^{3T} - P'^{1T} \\ y_2 P'^{3T} - P'^{2T} \end{bmatrix} \tag{4}$$

Since X is homogeneous coordinates, which is only three independent degrees of freedom with scale, however, linear equation set $AX = 0$ contains four equations, thus the above system is over determined system. To find the approximate solution of the equation $AX = 0$, we can make it as a follow optimization:

$$\min_X ||AX|| \tag{5}$$

with $||AX|| = 1$.

Primarily, we obtain the initial value of X by the above formula, and then optimize the value of X using the method of multiple view projective reconstruction by bundle adjustment, we get the final X by minimizing the follow function:

$$\min \sum_{i=1}^{n} v_i d(Q(X, I_i), x_i)^2 \tag{6}$$

Where I_i is the i KeyFrame, if there have mapping point on the I_i image, we set $v_i = 1$, else set $v_i = 0$. $Q(X, I_i)$ is the projection on the i KeyFrame and the $d(Q(X, I_i), x_i)^2$is the reprojection errors computed using the corresponding projection matrices.

III. Experimental Results

In order to verify the superiority of our measurement technology, in our experiments, we select the DJI Phantom 3 Professional as our mobile platform and the built-in camera as monocular camera, we totally test three different video sequences captured using our monocular camera in our campus. Each video is captured by 40 frames per second with a resolution of 1280×720.

We carefully plan the average flight path for unmanned aerial vehicle (UAV), and manipulate the UAV(DJI) flying in the sky over the targets, to make it emerge the closed loop in its flight path as far as possible, which can aid to make full use of the function of the loop closing module in ORB-SLAM algorithm, the loop closing module can reduce the accumulative error effectively, which can improve the estimation accuracy of the KeyFrames. In order to solve the inherent scale uncertainty of the monocular camera, we place a calibration board in our scenario, which aids to calculate the scale of the scenario by the relative poses and true poses of the KeyFrames.

From relative poses to real: We totally test three apparently different scenes in our experiments including the indoor and outdoor cases. When we deal with the original video sequences of the scenes using the ORB-SLAM algorithm, we can estimate a set of relative KeyFrames poses. In order to get the true poses from the relative poses, we first use the calibration board in the scene to calculate the scale in advance, the results see the second column in Table I, then we successfully change the relative poses to real, part of the results see the third and fourth columns in Table I. Due to the rotation vector is independent of the scale factor, so we just list the translation vector in the Table I, as we can easily see in the Table I, the different scenarios have different scales but the same scenario has the same value, which yet verify the aforementioned conclusion, that the scales among all of the KeyFrames are unified value.

Tracking the target: To get the image coordinates of the target in all KeyFrame sequences, we make it as a tracking problem. As we mention in the previous sections, on the basis

TABLE I
SCALE, RELATIVE AND REAL POSES OF KEYFRAMES.

Scenes	Scales	Relative Poses				Real Poses			
		kf1	kf2	kf3	kf4	kf1	kf2	kf3	kf4
I	**3.75**	-0.381075 -0.060165 0.0847729	-0.556926 -0.116769 0.151738	-1.040679 -0.237356 0.551877	-1.064672 -0.283631 0.611086	-1.516709 -0.229499 0.2674623	-2.987029 -0.409280 0.7776077	-3.903395 -0.892175 1.996418	-3.976020 -1.109835 2.222299
II	**3.95**	-0.534567 -0.039055 0.251774	-0.413442 -0.016587 0.215817	0.231164 -0.006011 0.102690	0.257555 -0.004890 0.086608	-2.141416 -0.146207 0.990824	-1.643556 -0.121913 0.798639	0.913986 -0.041968 0.357104	1.042089 -0.031596 0.322696
III	**8.90**	0.043163 0.272785 -0.042908	-0.169926 0.014856 -0.035846	0.068050 -0.297513 -0.000541	0.077502 0.210155 -0.086869	-1.814665 1.713910 -0.300230	-3.503213 -1.338974 0.003736	-1.318560 -4.715826 0.424791	0.691467 0.957453 -0.677240

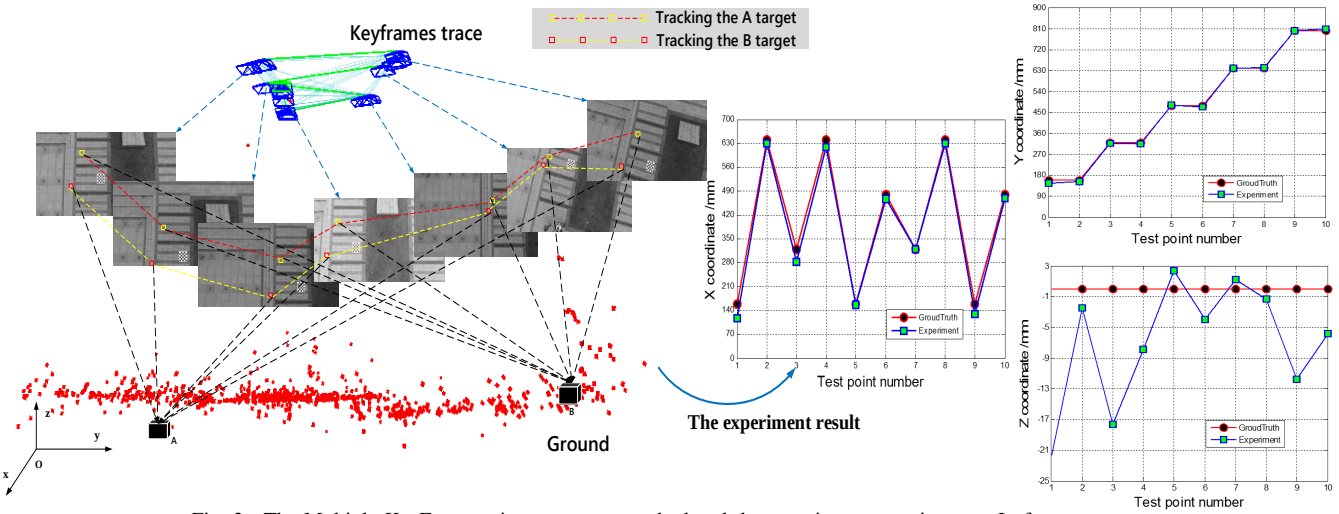

Fig. 3. The Multiple KeyFrames view geometry method and the experiment error in scene I of our system.

TABLE II
THE 3D GEODETIC COORDINATE RESULTS IN OUR EXPERIMENT.

Scenes	Experimental values	Ground true
I	[0.6299, 0.1539, -0.002] [0.6197, 0.3146, -0.007] [0.1563, 0.4807, 0.0023] [0.3202, 0.6413, 0.0012] [1.7613, 2.3836, -0.001]	[0.6400, 0.1600, 0.0000] [0.6400, 0.3200, 0.0000] [0.1600, 0.4800, 0.0000] [0.3200, 0.6400, 0.0000] [1.8000, 2.4000, 0.0000]
II	[0.1897, 0.1496, 0.0215] [0.6625, 0.1479, 0.0195] [0.3398, 0.3024, 0.0160] [0.6500, 0.6237, 0.0141] [0.4986, 0.7813, 0.0133]	[0.1600, 0.1600, 0.0000] [0.6400, 0.1600, 0.0000] [0.3200, 0.2300, 0.0000] [0.6400, 0.6400, 0.0000] [0.4800, 0.8000, 0.0000]
III	[0.3168, 0.0962, 1.1042] [0.1794, 0.3485, 0.9832] [0.0582, 0.4781, 1.0234] [0.3055, 0.4622, 0.9827] [0.3078, 0.5960, 1.2315]	[0.6400, 0.1600, 0.0000] [0.1600, 0.3200, 0.0000] [0.1600, 0.4800, 0.0000] [0.3200, 0.4800, 0.0000] [0.0111, 0.8000, 0.0000]

of the latest ORB-SLAM algorithm [14], we have designed a lean ORB-SLAM system added an independent tracking module which purpose is to track the target in the original video sequences, and then we record the image coordinates of the target whenever a new KeyFrame inserts.

In our experiment, we adopt one of the most successfully tracking algorithm STC to deal with the problem, which has good tracking effect to apparent object. In fact, in our indoor experiments, it works well and give us the accurate image coordinates, but there is not very good tracking performance in outdoor scenarios, the reason is that the target becomes small in outdoor filming scenario and it is hard to detected, so in our future work, we will mainly focus on the small targets tracking under the outdoor scenes.

Calculation and optimization: From the two modules above, we have obtained the accurate projection matrix of the all KeyFrames and its corresponding image coordinates of the target. There are two steps when we calculate the 3D geodetic coordinate, the first module is to get the initial value of the target using the double view orientation method, and the second module is to make further iteration optimization using multiple views projective reconstruction method see Figure 3 by bundle adjustment. Parts of the results show in Table II.

IV. CONCLUSION

In this work, we have presented a accurate 3D geodetic coordinate calculating method based on monocular vision on UAV platform, and with a detailed description of its framework and process, which include camera pose estimation with ORB-SLAM, tracking the target in the KeyFrame sequences, calculation and optimization. Compared with the other existing measurement techniques, our system has two outstanding advantages, for one thing, our system not only has the simple structure and operation, but also the flexible camera view. For another, based on our experiments with different scenes from indoor to outdoor, the accuracy of our system is typically

below 1 cm in small indoor scenarios and of a few meters in large outdoor scenarios.

The most important innovation in our system is the virtual scalability that we cleverly make the multiple KeyFrame as different separate camera views, which is the basis of the whole system. And another key point is that we use multiple view projective construction method instead of the binocular positioning technology to calculate the 3D geodetic coordinate of the target, which can make full use of the information to make global optimization.

ACKNOWLEDGMENT

This work is supported by the ShenZhen Science and Technology Foundation (JCYJ20160229172932237), National Natural Science Foundation of China (No. 61672429, No. 61502364, No. 61272288, No. 61231016), Northwestern Polytechnical University (NPU) New AoXiang Star (No. G2015KY0301), Fundamental Research Funds for the Central Universities (No. 3102015AX007), NPU New People and Direction (No. 13GH014604).

REFERENCES

[1] D. M. Bevly, J. C. Gerdes, C. Wilson, and G. Zhang, "The use of gps based velocity measurements for improved vehicle state estimation," in *American Control Conference, 2000. Proceedings of the 2000*, vol. 4. IEEE, 2000, pp. 2538–2542.

[2] W. Qilong, L. Jianyong, S. Haikuo, S. Tengteng, and M. Yanxuan, "Research of multi-sensor data fusion based on binocular vision sensor and laser range sensor." *Key Engineering Materials*, vol. 693, 2016.

[3] Y. Cui, F. Zhou, Y. Wang, L. Liu, and H. Gao, "Precise calibration of binocular vision system used for vision measurement," *Optics express*, vol. 22, no. 8, pp. 9134–9149, 2014.

[4] D.-y. Ge, X.-f. Yao, and Z.-t. Lian, "Binocular vision calibration and 3d re-construction with an orthogonal learning neural network," *Multimedia Tools and Applications*, pp. 1–16, 2015.

[5] K.-S. Kwon, M.-H. Jang, H. Y. Park, and H.-S. Ko, "An inkjet vision measurement technique for high-frequency jetting," *Review of Scientific Instruments*, vol. 85, no. 6, p. 065101, 2014.

[6] H. Li, Y.-L. Chen, T. Chang, X. Wu, Y. Ou, and Y. Xu, "Binocular vision positioning for robot grasping," in *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1522–1527.

[7] Z. QIAN, X.-y. PENG, S.-l. JIA, and H. LIU, "Study on binocular vision positioning calibration algorithm of moving platform [j]," *Computer Simulation*, vol. 10, p. 069, 2012.

[8] D. Schneider, Y.-S. Gloy, and D. Merhof, "Vision-based on-loom measurement of yarn densities in woven fabrics," *Instrumentation and Measurement, IEEE Transactions on*, vol. 64, no. 4, pp. 1063–1074, 2015.

[9] T.-d. Tan and Z.-m. Guo, "Research of hand positioning and gesture recognition based on binocular vision," in *VR Innovation (ISVRI), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 311–315.

[10] Y. Tian, X. Chen, Q. Huang, P. Lv, R. Li, and M. Li, "3d localization of moving object by high-speed four-camera vision system," in *Future Intelligent Information Systems*. Springer, 2011, pp. 425–434.

[11] Y. XIA, Z. SU, X.-b. WU, Y. XIA, Z. SU, X.-b. WU, Y. XIA, Z. SU, and X.-b. WU, "Calibration of binocular vision system and its application in 3d measurement [j]," *Journal of Image and Graphics*, vol. 7, p. 014, 2008.

[12] Z. Jing, F. Z. Xue, and L. I. Zu-Shu, "Space target location measurement based on monocular vision," *Transducer Microsystem Technologies*, 2011.

[13] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 834–849.

[14] R. Mur-Artal, J. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *Robotics, IEEE Transactions on*, vol. 31, no. 5, pp. 1147–1163, 2015.

[15] O. G. Grasa, E. Bernal, S. Casado, I. Gil, and J. Montiel, "Visual slam for handheld monocular endoscope," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 1, pp. 135–146, 2014.

[16] X. He, Z. Cai, D. Huang, and Y. Wang, "Indoor navigation for aerial vehicle using monocular visual slam," in *CGNCC*, 2014, pp. 2071–2075.

[17] R. Munguia and A. Gra, "Monocular slam for visual odometry," in *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1–6.

[18] L. Ruotsalainen, S. Grohn, M. Kirkko-Jaakkola, L. Chen, R. Guinness, and H. Kuusniemi, "Monocular visual slam for tactical situational awareness," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*. IEEE, 2015, pp. 1–9.

[19] J. Ventura, C. Arth, G. Reitmayr, and D. Schmalstieg, "Global localization from monocular slam on a mobile phone," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 4, pp. 531–539, 2014.

[20] C. Wang, T. Wang, J. Liang, Y. Chen, Y. Zhang, and C. Wang, "Monocular visual slam for small uavs in gps-denied environments," in *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*. IEEE, 2012, pp. 896–901.

[21] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *Robotics, IEEE Transactions on*, vol. 28, no. 5, pp. 1188–1197, 2012.

[22] R. Mur-Artal and J. D. Tardós, "Fast relocalisation and loop closing in keyframe-based slam," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 846–853.

[23] R. Mur-Artal and J. D. Tardos, "Orb-slam: tracking and mapping recognizable features," in *Proceeding of Robotics: Science and Systems (RSS) Workshop on Multi View Geometry in Robotics*, 2014.

[24] R. Mur-Artal and J. D. Tardós, "Probabilistic semi-dense mapping from highly accurate feature-based monocular slam," *Proceedings of Robotics: Science and Systems, Rome, Italy*, 2015.

[25] L. Zhao, S. Huang, and G. Dissanayake, "Linear monoslam: A linear approach to large-scale monocular slam problems," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1517–1523.

[26] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski, "Bundle adjustment in the large," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 29–42.

[27] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.