Active Infrared Coded Target Design and Pose Estimation for Multiple Objects

Xudong Yan[1], Heng Deng[1], Quan Quan[2]

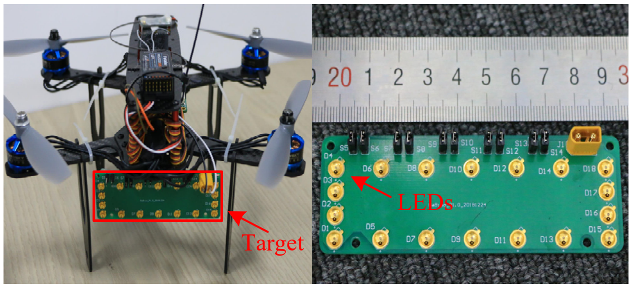
***Abstract*— Relative pose estimation is critical for collaborative multi-agent systems. To achieve accurate and low-cost localization in cluttered and GPS-denied environments, we propose a novel relative pose estimation system based on a designed active infrared coded target. Specifically, each agent is equipped with a forward-looking monocular camera and a unique infrared coded target. The target with the unique lighted LED arrangement is detected by the camera and processed with an efficient decoding algorithm. The relative pose between the agent and the camera is estimated by combining a PnP algorithm and a Kalman filter. Various experiments are performed to show that the proposed pose estimation system is accurate, robust and efficient in cluttered and GPS-denied environments.**

I. INTRODUCTION

In recent years, there is a booming interest in collaborative multi-agent systems [1], [2], [3]. Compared to an individual agent, multi-agent systems have the advantage over faster task completion, more robust to sensor failures and higher-precision pose estimation through sensor fusion. Multi-agent systems have numerous indoor applications such as cooperative surveillance, monitoring, search and rescue missions.

In multi-agent systems, precise knowledge of the relative location among swarms of agents is crucial for the success of collaborative tasks. Although the Global Positioning System (GPS) is available for localization, the multi-agent systems may work in a GPS-denied environment such as indoors or urban areas with high buildings. Thus, vision-based relative pose estimator has emerged to be one of the most popular solutions for multi-agent systems.

Nowadays, the model-based relative localization, which mainly relies on artificial landmarks, has become a common approach for multi-agent systems. ARTags [4] and AprilTags [5] were commonly used for mutual localization. They provided not only a pose estimation but also a unique ID for each agent. But those artificial landmarks required a large and flat area, which made them unsuitable for micro-aerial agents. Passive landmarks such as colored markers were employed to estimate the pose in some researches [6], [7]. The distinction among the UAVs based on different colors of markers has also been proved to be an effective method for multi-target identification [8]. However, the high dependency on the illumination condition for passive markers may decrease the accuracy. Some researches distinct target ID with LEDs blink at different frequencies. [9] used a Dynamic Vision Sensor (DVS) for pose estimation. A DVS can track frequencies up to several kilohertz. Therefore, pose estimation can be performed with low latency. However, low sensor resolution (128×128 pixels) limited precision. [10] used CMOS camera to obtain high-quality images with high latency and low sampling frequency compared to the DVS but suffered from motion blur. Simultaneous Localization and Mapping (SLAM) or Monte Carlo Localization (MCL) techniques were utilized to estimate positions for agents, as in [11]. However, when sharing the positions with each agent, the techniques were time-consuming.



(a) (b)

Fig. 1. Illustration of the hardware platform. (a) The agent with an infrared coded target on the back; (b) The designed target, which is 9.1 cm×2.2 cm with 18 infrared LEDs.

The purpose of this paper is to propose an efficient relative localization system with the designed coded target. The main contributions of this paper are as follows:

- Design a small active infrared coded target for extracting IDs for multiple micro-agents.
- Design a target decoding algorithm which is highly efficient and robust in cluttered environments.
- Estimate relative positions for multi-agents with lighted LEDs on the designed targets.
- Perform various experiments to evaluate the proposed pose estimated system.

The rest of the paper is organized as follows. In Section II, we describe the design of the coded target. The target decoding algorithm and the pose estimation algorithm are discussed in Section III. In Section IV, we present the experiments, and the concluding remarks can be found in Section V.

II. ACTIVE INFRARED CODED TARGET DESIGN

The designed hardware platform, as shown in Fig. 1(a), is composed of a small-sized coded target and a monocular camera. The coded target is a rectangle formed with 18 active

[1] Xudong Yan, and Heng Deng are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, P. R. China {yanxd, dengheng}@buaa.edu.cn

[2] Quan Quan is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, P. R. China, and also with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, P. R. China qq_buaa}@buaa.edu.cn

infrared LEDs, described in Fig. 1(b). The coded target is small enough to be mounted on micro-agents.

The designed target is presented in Fig. 2. Infrared LEDs on the target are divided into locators and detectors according to different functions, as shown in Fig. 2(a). The locators distinguish the target from the others while the detectors provide unique IDs. For classifying locators and detectors, their contour aspect ratios are designed to be different. As for locators, light spots of four adjacent infrared LEDs overlap together, making an aspect ratio equal to 3:1 approximately. On the other hand, the contour aspect ratio for detectors is about 1:1 since they are relatively sparse.

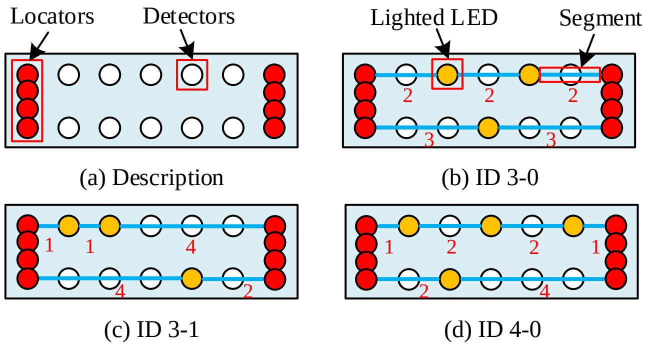
To extract a specific ID from the coded target, two constraints are designed, as shown in Fig. 2(b)(c)(d). One is the number of lighted detectors, defined as N. The other is segment ratio, defined as M. IDs in Fig. 2(b)(c) have the same number of lighted detectors ($N = 3$), but different segment ratios ([2:2:2:3:3] in Fig. 2(b) yet [1:1:4:4:2] in Fig. 2(c)); ID in Fig. 2(d) has different number of lighted detectors ($N = 4$) and different segment ratios ([1:2:2:1:2:4]) as well. Before extracting IDs, it is necessary to build an ID library. Inspired by modified lexicode[1] and AprilTag [5] which use minimum Hamming distance[2], we generate an ID library by maximizing Mean Squared Error (MSE):

$$p = \frac{1}{m}\sum_{i=1}^{m}(x_i - y_i)^2 \tag{1}$$

where x_i and y_i are corresponding segments from two IDs whose N is equal; m is the number of segments. When p is greater than a threshold, the ID is considered to be a valid candidate. MSE and minimum Hamming distance have a common goal, i.e., maximizing the difference between IDs. However, the proposed coded target is small and the number of detectors is limited. Therefore, using segment ratios to be a constraint is better than using a minimum Hamming distance, which is more robust as the number of detectors increases.

Moreover, the coding scheme should be robust to the rotation. Thus, when a target is rotated by 90, 180 and 270 degrees, the same ID should be extracted rather than the others. Since the proposed coded target is a rectangle instead of a square, the only situation to be considered is 180 degrees rotation. To solve this problem, another constraint is added: candidate should never be of central symmetry. With this constraint, the robustness against rotation is ensured in ID extraction.

In this paper, we set N ranging from 3 to 8. For every case of a certain number of N, we set 4 possibilities of the segment ratios. As a result, a total of 24 candidate IDs exists, which are enough for multi-agent systems. There are several advantages to the designed target. First, the proposed target is small enough to mount on micro-agents. Secondly, unlike passive landmarks, active infrared LEDs work under bad conditions regardless of dark or visible light pollution. Thirdly, the locators and the detectors are easy to distinguish and hard to be disturbed by other targets. Finally, as discussed later, the correspondence between LEDs and contours in the camera image is highly efficient due to the design of our targets.



(a) Description

(b) ID 3-0

(c) ID 3-1

(d) ID 4-0

Fig. 2. Introduction of the designed target in details. (a) Locators (8 LEDs) are set on the left and right sides, colored in red and constantly lighted; detectors (10 LEDs) are set on the top and bottom sides, colored in white and whether the detectors are lighted depends on the coding algorithm, (b)(c)(d) represent different IDs with a different number of lighted detectors and segment ratio. Detectors colored in yellow are lighted and the segments are colored in blue. The red number represents segment ratio. Label "ID 3-0" means that the number of lighted detectors is 3 and the segment ratio is [2:2:2:3:3].

III. Pose Estimation for Multiple Objects

Each agent in the multi-agent system is equipped with the designed target described in Section II. The proposed ID extraction algorithm and the relative pose estimation method mainly contains five steps, as shown in Fig. 3. The current camera image, aspect ratio, LED configuration and ID library serve as inputs to the proposed algorithm. Meanwhile, the target ID, relative position ($\mathbf{x}_{ij}^c$) and relative orientation ($\mathbf{r}_{ij}^c$) act as outputs. First, lighted LEDs in the camera images are detected. Second, they are divided into two parts according to the aspect ratio, which are locators and detectors. Then, using locators to distinguish each target from the others in the camera images. Next, IDs are extracted with ID library. Finally, relative pose and orientation are estimated for the multi-agent. All steps are described below in more detail.

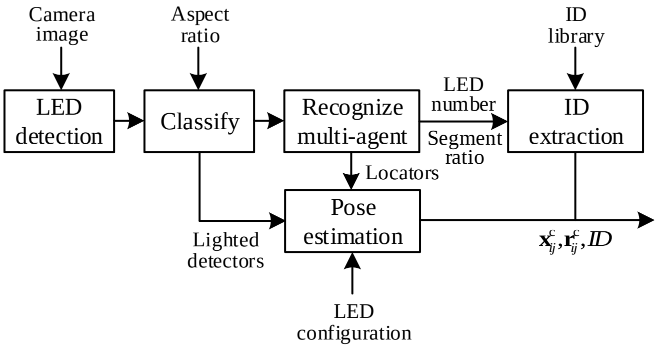


Fig. 3. Pipline of the ID extraction and relative pose estimation algorithm.

[1] https://en.wikipedia.org/wiki/Lexicographic_code

[2] https://en.wikipedia.org/wiki/Hamming_distance

A. LED Detection and Classify

First, LEDs need to be detected in the image. Since infrared LEDs' wavelength matches the infrared-pass filter in the camera, they appear very bright in the image compared to the background. Thus, a thresholding function is sufficient to detect LEDs,

$$I'(u, v) = \begin{cases} 255, & \text{if } I(u, v) > T, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $I(u,\ v)$ is the original pixel value and $I'(u,\ v)$ is the processed one. T is the threshold. We found that a large range of threshold parameter works well (80-180). Then, the contours of the bright area are calculated by findContours function. Next, the first moment is utilized to calculate the centers of the contours and minAreaRect function to calculate the minimum-area bounding rectangle for the contours. The above functions are achieved by the OpenCV library[3].

Afterward, the contours are classified as two parts, locators (L) and detectors (D_n), by the aspect ratio of the minimum-area bounding rectangle. When the threshold is set to 2, it can effectively classify the locators (aspect ratio nearly equal to 3:1) and the detectors (aspect ratio nearly equals to 1:1).

B. Recognize multi-agent

Before extracting IDs, the targets in the camera image need to be separated first, with the help of the locators. First, the locators are sorted by x-axis in the image to search pairs of locators (L_1, L_2). Since adjacent locators are more likely to form the target, sorting the locators can reduce the algorithm complexity greatly. After picking a pair of locators L_1 and L_2, the shape formed by L_1 and L_2 is to be determined whether a target or not by geometric relationship. Specifically, it is more likely to be a target when the shape is a rectangle. Thus, two constraints are used: the opposite side to be equal and angles between adjacent sides close to 90 degrees. The first constraint is to ensure the shape formed by L_1 and L_2 to be a parallelogram, avoiding the case that shape is formed by different targets' locators with different orientations, as shown in Fig. 4(a). The distance constraint is described as follows,

$$c_1 = \sqrt{(a-b)^2 + (c-d)^2} < T_d \tag{3}$$

where a, b, c, d are the side lengths of the shape, as shown in Fig. 4(a) and T_d is the threshold.

The second constraint is to ensure the angles between adjacent sides close to 90 degrees, avoiding the case that two locators result from different targets with same orientation but different positions, as shown in Fig. 4(b). We use Cosine theorem to restrict the angles,

$$c_2 = \frac{\mathbf{v_1} \cdot \mathbf{v_2}}{\|\mathbf{v_1}\| \|\mathbf{v_2}\|} < T_a \tag{4}$$

where $\mathbf{v_1} \in \mathbb{R}^2$ and $\mathbf{v_2} \in \mathbb{R}^2$ are the direction vectors of the adjacent sides and T_a is the threshold. When the two constraints are satisfied, the shape formed by L_1 and L_2 are determined to be a target.

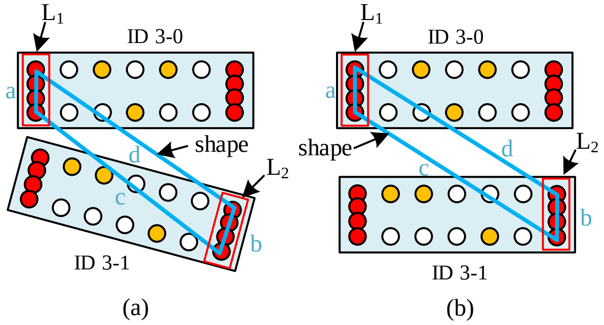


Fig. 4. Two situations when picking a pair of locators to form the shape. (a) locators picked from different targets with different orientations. a, b, c, d represent the sides of the shape. (b) locators picked from different targets with same orientation but different positions.

C. ID Extraction

After separating the target from the others, the ID can be extracted through the proposed ID extraction algorithm. First of all, endpoints of the target need to be calculated for computing the top edge (TE) and the bottom edge (BE) of the target. The endpoints can be calculated by the geometric characteristics of the locators since the endpoints are the locators' five equal points. Then, compute distances between lighted detectors and TE, BE to determine which line detectors belong to. Next, the segment ratio can be computed, as shown in Fig. 2(b)(c)(d). Finally, the ID can be extracted using (1) with ID library. The method of recognizing multi-agent and ID extraction is summarized in Algorithm 1.

Algorithm 1 Recogizing multi-agent and ID extraction

```
Input:
    Locators $L$ and detectors $D_n$ in a image
Output:
    IDs of the targets in the image
Sort $L$ by x-axis in the image
for $L_1 \in L$ do
    for $L_2 \in L \setminus L_1$ do
        if $c_1 < T_d$ and $c_2 < T_a$
            Compute $TE$, $BE$
            Compute distances between $D_n$ and $TE$, $BE$
            Compute segment ratio
            Extract ID using MSE with ID library
        end if
        $L \leftarrow$ update($L$)
    end for
end for
```

D. Relative Pose Estimation

The designed coded target can also be used for estimating relative pose. As shown in Fig. 3, the inputs of the pose estimation step are locators, lighted detectors and LED configuration. Locators and lighted detectors have been obtained in the above steps. The LED configuration, i.e., the positions of the lighted LEDs in the reference frame of the target, needs to be measured in advance instead. The advantage of the proposed target is that matching the lighted LEDs on the

[3] https://opencv.org/

target with the contours in the camera image is easy when the target is separated from the others. Compared with previous methods such as correspondence search in [12] and Particle Swarm Optimization (PSO) algorithm in [13] to find the optimal correspondence, our method is linear. The relative pose is initialized by the PnP algorithm first. The coordinates of the three lighted LEDs are fed into the P3P algorithm, as described in [14]. Then, other lighted LEDs on the target can be used to refine the reprojection error and evaluate its correctness. After the initialization based on PnP algorithm, a Kalman filter is utilized to predict relative pose to improve the accuracy and robustness, as described in [15], with real-time operation state estimation.

IV. Experiments and Results

We design a set of experiments to evaluate the accuracy, robustness and efficiency of our system. In Experiment 1, a camera is used as a static agent who is fixed with a tripod, and one unmanned aerial vehicle (UAV) equipped with the designed target is involved in position estimation. In Experiment 2, two more UAVs are added to the pose estimation progress. This experiment aims at applying the proposed system into multi-agent situations. Occlusion experiments are designed to evaluate the robustness of the system as well. Finally, the computational cost is measured for each step of the algorithm to evaluate the efficiency of the system in Experiment 3.

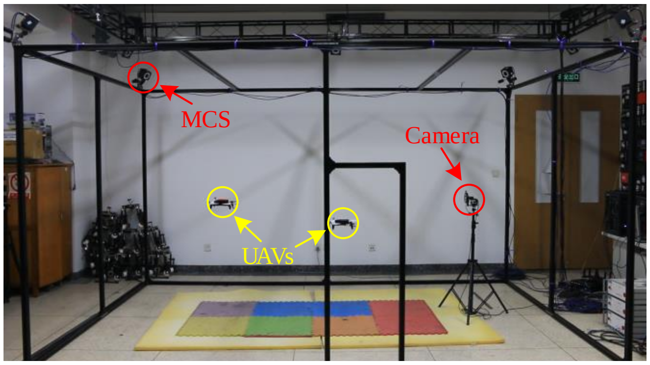


Fig. 5. Indoor experiment environment. MCS is used for providing the ground truth. The camera is fixed with a tripod and the UAVs are moving.

A. *Experiment 1: Single-agent localization*

The goal of this experiment is to characterize the relative localization system and analyze its accuracy. A camera with an infrared-pass filter, a resolution of 752 × 480 pixels and a field of view of 90° is used for the experiments. The coded target is equipped on the back of the UAV (see Fig. 1(a)). As shown in Fig. 5, the experiment is carried out in a laboratory room with an OptiTrack[4] motion capture system installed, which provides the ground truth of the UAV's positions and orientations.

In the first run, the camera is positioned at a fixed location while the UAV is moving. The proposed relative localization system obtains UAV's position and orientation relative to the camera. The estimation results are then transformed into the world frame compared with the ground truth provided by OptiTrack motion capture system. In Fig. 6, the estimated position, errors of position and orientation of the UAV are shown in (a), (b), and (c), respectively. Accuracy assessment results are as follows: with a total of 3978 images in the video dataset, there are 3 images (0.075%) that the target could not be detected while other images (99.925%) found a good pose estimation; the position error is between 0.05 cm and 17.98 cm with a mean of 2.42 cm and a standard deviation of 1.91 cm; the orientation error lies between 0.05° and 2.09° with a mean of 0.3° and standard deviation of 0.21°.

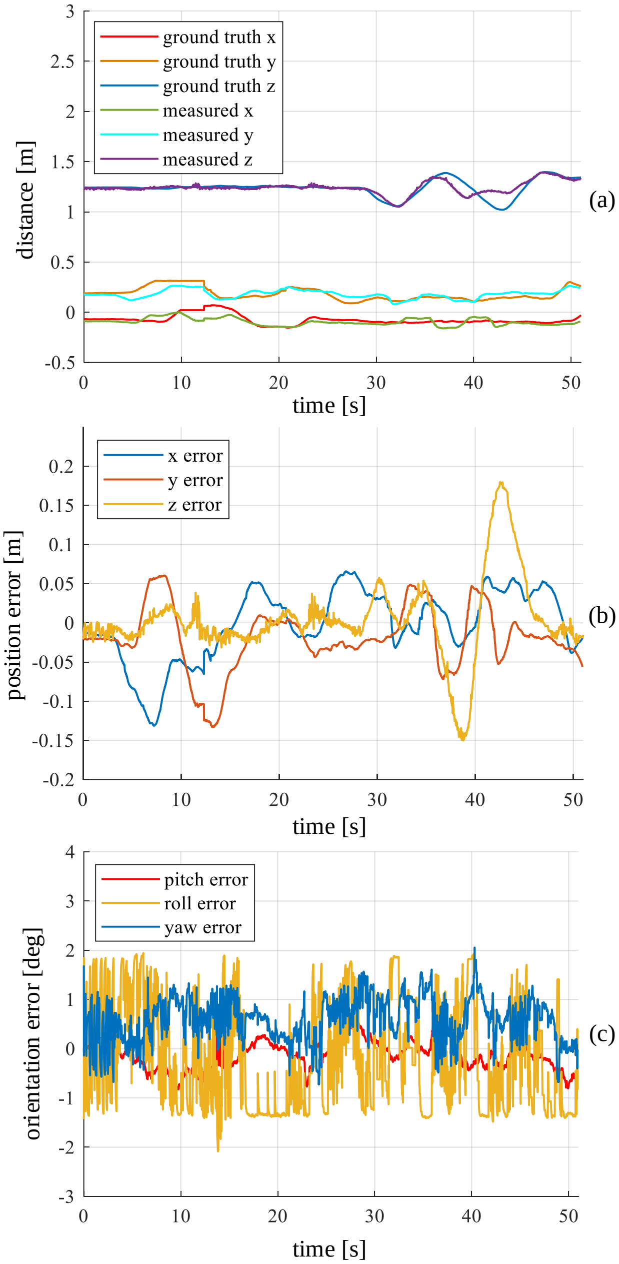


Fig. 6. The performance of single-agent localization, including the estimated position, error of position and error of orientation.

[4] https://optitrack.com/

In the second run, we evaluate the error concerning the distance between the camera and the UAV. The camera was fixed at the origin and the target is moved from 0.3 m to 2.4 m while recording a total of 3759 images. Fig. 7 shows the boxplot of the position error. We could observe that only a minor increase exists in the localization error when the distance between the camera and the target increases to the maximum range. However, the target size and the camera resolution limit the maximum valid distance between the camera and the coded target. At a distance of 2.5 m, the size of the detected LEDs is reduced to only a few pixels in the camera of 752 × 480 resolution and the luminous area emitted by LEDs overlapped.

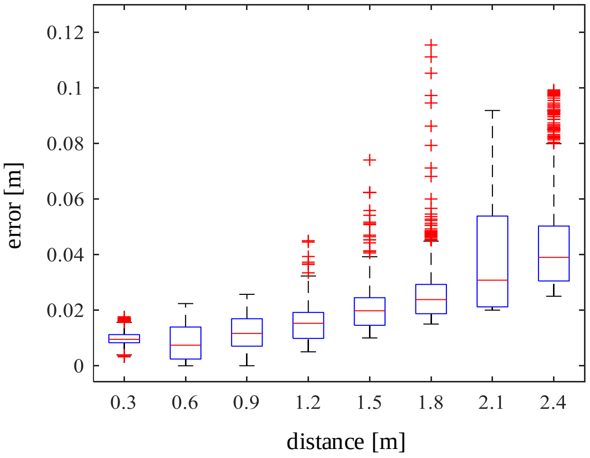


Fig. 7. Boxplot of the position error concerning the distance between the camera and the target.

B. Experiment 2: Multi-agent recognition and localization

The second experiment firstly aims at the application of pose estimation for multi-agent. The experimental setup is the same as Experiment 1 except for increasing two UAVs equipped with the coded targets. The camera is positioned at a fixed location while three UAVs are moving at the same time. The estimated pose of one of the UAVs is recorded to evaluate the proposed system. In Fig. 8, the estimated position, errors of position and orientation of this UAV are shown in (a), (b), and (c), respectively. Accuracy assessment results in the multi-agent system and occlusion situations are as follows: the position error is between 0.09 cm and 19.86 cm with a mean of 3.01 cm and standard deviation of 2.10 cm; the orientation error lies between 0.08° and 12.74° with a mean of 1.43° and a standard deviation of 1.08°.

Secondly, the designed occlusion experiments are performed to evaluate the robustness of the system, as shown in Fig. 9. In the T_1 stage, the three targets are not obstructed from each other. Their positions, orientations, as well as IDs, can be estimated by the proposed algorithm, as shown in Fig. 9(a). In the T_2 stage, the locators of the target in the middle is occluded by the right target. Therefore, the middle target is not recognized by the proposed algorithm, but the right one does not get disturbed. In the next stage, the left one and the right one all occlude the middle target, but they do not get disturbed by the detectors of the middle one, as described in Fig. 9(c). Finally, the middle target appears and is recognized again by the proposed algorithm. The occlusion experiments show that when the targets' locators are occluded, the target is hard to be recognized but the remaining detectors do not interference other targets. And once the target appears, it will be recognized immediately.

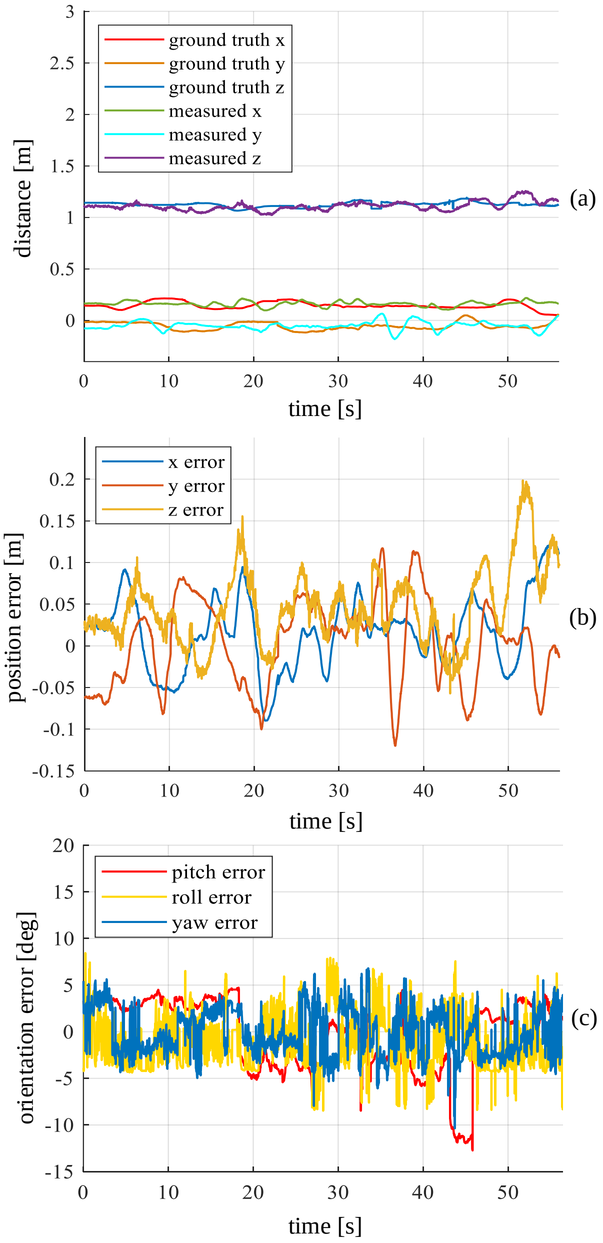
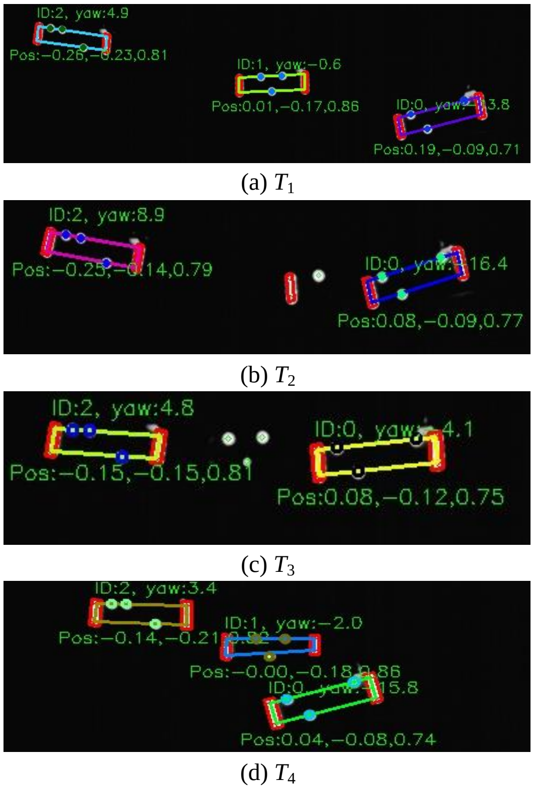


Fig. 8. When there are 3 targets observed at the same time, the estimated position, position error and orientation error of one of them are recorded as (a), (b), (c).

C. Experiment 3: Execution times

The mean execution times for each step of the proposed algorithm with a different number of targets can be found in Table I. For every fixed number of targets, they are measured on a dataset with 2600 images. We use a laptop with an Intel



(a) T_1

(b) T_2

(c) T_3

(d) T_4

Fig. 9. Camera images (a) before, (b)(c) during, (d) after the targets occlusion. The green text in the images shows the ID of the targets, yaw and position relative to the camera.

i7-7700 (2.80 GHz) processor. Note that as the number of targets increases, the execution times increases sublinearly. On average, relative pose estimation makes up 43.5% of the execution time. And thanks to the designed coded target, ID extraction step is efficient enough, which makes up only 7.72% of the execution time.

TABLE I

EXECUTION TIMES

Process	Running Time (ms)			
	1 target	2 targets	4 targets	8 targets
LED detection	0.802	0.807	0.887	1.016
Classify	0.555	0.692	1.298	1.988
Recognition	0.582	0.794	1.660	2.452
ID extraction	0.229	0.311	0.645	1.276
PnP + KF	1.225	1.806	4.039	6.657
Total	3.393	4.410	8.529	13.39

V. CONCLUSIONS

In this paper, we present an accurate, robust and efficient system capable of providing relative pose estimation and target ID for multi-agent systems based on the designed infrared coded target. The designed target is simple and only requires a few infrared LEDs to code unique IDs and estimate pose. As demonstrated, our solution works well in cluttered environments for micro multi-agent systems, further, owns high accuracy for both ID extraction and pose estimation. Experiment results also show that, compared with previous methods, the new strategy is more efficient because of the easy correspondence method and the execution time is sublinear with targets increased.

In the future, we plan to open source the code for everyone to integrate our target into their robotic platforms to solve the problem of multi-agent system relative localization. Meanwhile, we will continue to improve the accuracy and efficiency of the system and try to solve the problem that the maximum valid distance between the camera and the coded target is limited.

REFERENCES

[1] Q. Quan, Introduction to Multicopter Design and Control. Singapore: Springer, 2017.

[2] Y. Stergiopoulos, M. Thanou, and A. Tzes, “Distributed collaborative coverage-control schemes for non-convex domains,” in *IEEE Transactions on Automatic Control*, vol. 60, no. 9, pp. 2422-2427, Sept. 2015.

[3] Y. Kantaros, M. Thanou, and A. Tzes, “Distributed coverage control for concave areas by a heterogeneous Robot-Swarm with visibility sensing constraints”, in *Automatica*, vol. 53, pp. 195-207, Mar. 2015.

[4] M. Fiala, “ARTag, a fiducial marker system using digital techniques,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, 2005, pp. 590-596.

[5] E. Olson, “AprilTag: A robust and flexible visual fiducial system,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, 2011, pp. 3400-3407.

[6] S. Roelofsen, D. Gillet, and A. Martinoli, “Reciprocal collision avoidance for quadrotors using on-board visual detection,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, 2015, pp. 4810-4817.

[7] A. Franchi, C. Masone, V. Grabe, M. Ryll, H. Bulthoff, and P. Giordano, “Modeling and control of UAV bearing formations with bilateral highlevel steering,” in *International Journal of Robotics Research*, vol. 31, no. 12, pp. 1504-1525, 2012.

[8] R. Tron, J. Thomas, G. Loianno, J. Polin, V. Kumar and K. Daniilidis, “Vision-based formation control of aerial vehicles,” in *Robotics: Science and Systems*, 2014.

[9] A. Censi, J. Strubel, C. Brandli, T. Delbruck, and D. Scaramuzza, “Low-latency localization by active LED markers tracking using a dynamic vision sensor,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, 2013, pp. 891-898.

[10] D. Dias, R. Ventura, P. Lima and A. Martinoli, “On-board vision-based 3D relative localization system for multiple quadrotors,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 2016, pp. 1181-1187.

[11] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, “Vision-based state estimation and trajectory control towards high-speed flight with a quadrotor,” in *Robotics: Science and Systems*, 2013.

[12] M. Faessler, E. Mueggler, K. Schwabe and D. Scaramuzza, “A monocular pose estimation system based on infrared LEDs,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, 2014, pp. 907-913.

[13] W. Su, A. Ravankar, A. A. Ravankar, Y. Kobayashi and T. Emaru, “UAV pose estimation using IR and RGB cameras,” in *IEEE/SICE International Symposium on System Integration (SII)*, Taipei, 2017, pp. 151-156.

[14] L. Kneip, D. Scaramuzza, and R. Siegwart, “A novel parametrization of the perspective-three-point problem for a direct compuation of absolute camera position and orientation,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, 2011, pp. 2969-2976.

[15] Y. Xie, F. Pan, B. Xing, Q. Gao, X. Feng and W. Li, “A new on-board UAV pose estimation system based on monocular camera,” in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, 2016, pp. 504-508.