Vision-Aided Multi-UAV Autonomous Flocking in GPS-Denied Environment

Yazhe Tang, Yuchao Hu, Jinqiang Cui, Fang Liao, Mingjie Lao, Feng Lin, and Rodney S. H. Teo

***Abstract*—This paper presents a sophisticated vision-aided flocking system for unmanned aerial vehicles (UAVs), which is able to operate in GPS-denied unknown environments for exploring and searching missions, and also able to adopt two types of vision sensors, day and thermal cameras, to measure relative motion between UAVs in different lighting conditions without using wireless communication. In order to realize robust vision-aided flocking, an integrated framework of tracking-learning-detection on the basis of multifeature coded correlation filter has been developed. To achieve long-term tracking, a redetector is trained online to adaptively reinitialize target for global sensing. An advanced flocking strategy is developed to address the autonomous multi-UAVs' cooperative flight. Light detection and ranging (LiDAR)-based navigation modules are developed for autonomous localization, mapping, and obstacle avoidance. Flight experiments of a team of UAVs have been conducted to verify the performance of this flocking system in a GPS-denied environment. The extensive experiments validate the robustness of the proposed vision algorithms in challenging scenarios.**

***Index Terms*—Flocking, unmanned system, visual sensing.**

I. Introduction

THIS paper is inspired by flocking behaviors in nature via vision exploring in obstacle-rich environments. Different from conventional systems, this paper adopts a day camera and a thermal camera to conduct flocking using visual sensing in various illumination with autonomous navigation capabilities. Naturally, flocking performs a collective behavior that individuals use limited environmental information organizing in an ordered motion, so that they remain together as a group. This phenomenon widely exists in the nature [1] exhibited by many living beings such as fish, insects, and birds.

Flocking behavior was first simulated on a computer in 1987 by Reynolds [2] with an artificial life program via three heuristic rules, called boids, which only considers obstacle-free environments. Many variations of these rules considering obstacle avoidance and goal seeking have been investigated during these years. In addition, Graph theory has been utilized on alignment topics [3], artificial potentials [4] for obstacle avoidance as well as leader–follower schemes for goal searching [5].

Current flocking strategies can be roughly divided into two classes: centralized and decentralized. Centralized strategies use a central unit to organize behaviors of flocking systems [6]. They may become technically unfeasible for large-scale systems. Decentralized strategies are built based on the interactions between unmanned aerial vehicles (UAVs), which are mostly inspired by the evidence of biological systems or natural phenomena. As a typical decentralized strategy, the leader–follower-based method [7] considers some team members as leaders, and some act as followers. But each follower should have at least one leader, and some team members may act as both leaders and followers. This hierarchical architecture can simplify the flocking control to an individual motion planning. In a string or chain structure, each UAV follows one leader which is ahead of it [8]. The advantages of leader–follower methods are that the number of UAVs in a flocking group is scalable and the communication load is low.

To realize the flocking in radio silent situation, we propose to use passive vision sensing to achieve relative distance measurement in flocking. Vision sensing is widely used in autonomous systems due to the low cost of camera sensors and the rich information. Zhao *et al.* [9] presented a vision-based autonomous cargo transfer system using a UAV platform. They employed an on-board camera to detect and track the ellipse for task operation. One drawback is that it requires a standard shape as an auxiliary landmark for object detection and it is hard to be extended to general uncontrolled natural environments. Mueller *et al.* [10] designed an aerial tracking system for a UAV to track a ground moving target. A relay tracking scheme is designed based on a handover strategy for persistent tracking. The experiment shows that the persistent tracking is achievable only in simple ground background, and cluttered background may degrade its performance. In [11], tracking-learning-detection (TLD) [12] is used for target tracking from a UAV and achieves a good result in the short term. Their method only tracks a person at a height of 1.5 m but does not address common tracking problems such as large perspective change and scale variation. In addition, a tracker is proposed in [13] on the basis of multiple instance learning (MIL) [14] to track the aircraft/intruder in the sky from a fixed-wing UAV. They present a model-free aircraft tracking that can operate in an illumination changing environment. However, the tracking background is a clean sky, which is very ideal for target tracking. Recently, a tracking algorithm for

Manuscript received April 19, 2017; revised August 6, 2017 and November 13, 2017; accepted March 2, 2018. Date of publication April 9, 2018; date of current version August 31, 2018. *(Corresponding author: Jinqiang Cui.)*

The authors are with the Temasek Laboratories, National University of Singapore, Singapore 119077 (e-mail: yztang2008@yahoo.com; tslhuyu@nus.edu.sg; jinqiang@u.nus.edu; tsllf@nus.edu.sg; tsllaom@nus.edu.sg; linfeng@nus.edu.sg; tsltshr@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2824766

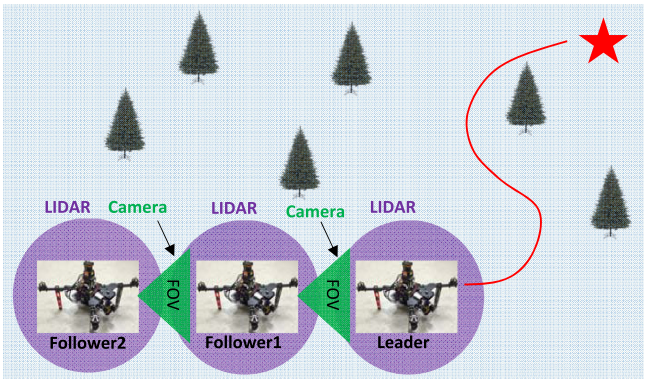


Fig. 1. Vision-aided flocking system exploring at unknown environment.

sense-and-avoid maneuvers by utilizing a combination of feature point tracking and morphological filters has been presented in [15].

The aforementioned papers present representative systems for vision-aided UAV tracking applications. Those vision sensing algorithms can be divided into model-based and model-free ones. The model-based tracking systems require auxiliary marker present for geometry mapping [16]. These methods can achieve accurate tracking results. However, markers may not be available in general environments. In contrast, model-free methods provide much more flexible and noninvasive solutions, which are very useful for practical applications. In this paper, we focus on a markerless vision sensing algorithm for concealment concern.

To achieve robust vision sensing for the proposed autonomous flocking system, we design a discriminative tracking algorithm in a closed-loop framework for tracking-learning-detection, which is applicable to both thermal and visible cameras for day and night operation. Taking into account the robust and real-time performance of the system, powerful and computational efficiency algorithms are essential for flocking. We employ a correlation filter to exploit multilevel-oriented gradient features for visual sensing. To achieve the long-term tracking, the correlation filter is trained online to adaptively reinitialize the target for the purpose of global sensing. The filter can wisely update the discriminative model based on the tracking quality, which may effectively prevent corruption of the tracker by the accumulative error. With a reliable visual feedback, a sequence of waypoints can be estimated for path planning implementation.

Fig. 1 illustrates the proposed vision-aided leader–follower flocking system for UAV cooperative flight, which is used as a core technology to aid a UAV searching system to maintain basic cooperative capability in GPS-denied and radio-silent environments. LiDAR sensor is equipped to provide point cloud information for the UAV to accurately sense the surrounding environment, which facilitates the obstacle avoidance and navigation for autonomous flights. In this system, the first UAV in the team serves as the team leader. It generates an online path from its current position to a given destination by using an improved A* search algorithm. The rest UAVs are called followers. The goal for each follower is selected from the trajectory points of its leader UAV that is just in front of it with a predefined safety distance away from the leader. For concealment concerns, communication has been minimized during the flocking, which is allowed only when the front UAV is completely occluded for a certain period. The system is expected to work in unknown environments with different lighting conditions. Therefore, both day and thermal cameras have been employed, and only the suited one will be used in accordance to environment illumination. To the best of our knowledge, this work is the first attempt to implement vision-aided autonomous UAV flocking in GPS-denied environments. Existing UAV flocking systems [17] normally adopt fixed-wing UAV flying at a high altitude with clean sky background. While our system operates at a low altitude in complicated environments, which implicitly increases the difficulty of tracking.

The remainder of this paper is organized as follows. Section II summarizes the system logic and flight control. Section III elaborates the correlation filter based vision sensing algorithm including tracking, detection, and learning scheme for the long-term visual tracking. Section IV introduces the autonomous navigation system. Section V presents real flight experiments to verify the proposed flocking system. Finally, we conclude this paper in Section VI.

II. Framework of the Flocking System

This paper depicts a search scenario that a flock of UAVs operates in an unknown and obstacle-rich environment with a tandem formation, which has three basic requirements: 1) each individual UAV can avoid obstacles; 2) maintain a team via vision of the follower's camera; and 3) reach a goal by following the team leader who knows the destination. Each UAV is an intelligent agent in an autonomous fashion for task operation. In the navigation module, LiDAR has been equipped for UAVs to sense the surrounding environment for navigation and obstacle avoidance. On-board cameras have been installed for the followers to track their leaders' waypoints that are employed to generate their trajectory reference.

In the flocking, the UAV with a given destination is named as a team leader and other UAVs are named as followers though they may be a corresponding leader of the following UAV. As shown in Fig. 1, a team with three UAVs has one leader which flies in the front of the team, and the rest of the two UAVs serve as followers in this team. For example, follower #1 follows the team leader's step and it also serves as the leader of follower #2 in the meantime. All the UAVs are equipped with an LiDAR sensor and able to navigate in unknown environments. Besides the LiDAR sensor, each follower UAV is equipped with an on-board camera, thus it is able to identify its corresponding leader for target following.

Due to limited communication among UAVs, the followers estimate their corresponding leaders' relative position by vision sensing. The camera module consists of a normal and a thermal camera. The working camera will be selected based on the ambient light condition, for instance, day and night. As the leader does not have any information about its followers, its acceleration has to be constrained, otherwise the followers will lag behind it and lose their leaders in each camera view, and

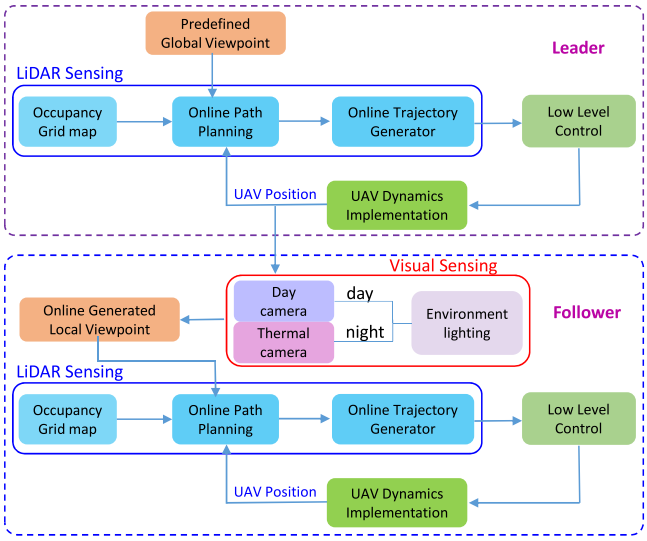


Fig. 2. Autonomous system structure with leader-and-follower logic of flocking UAVs.

will not be able to maintain the flocking. Moreover, since the leader may change direction and move away from followers in the time between two consecutive looks, the follower's heading is controlled such that it always points to the leader.

As illustrated in Fig. 2, the basic navigation and obstacle detection capabilities for each UAV have been provided by the aforementioned LiDAR-based simultaneous localization and mapping (SLAM). The leader and follower UAVs will generate high-level commands like trajectory following and vision-based target tracking, respectively. Based on the above navigation information, obstacle map and high-level commands, the path planning module will generate a collision-free path for low-level trajectory generation and flight control to realize autonomous flight.

During the flocking, the team leader runs a motion planning algorithm [18] to fly to a predefined destination along a collision-free trajectory. The followers estimate the relative position to their corresponding leader via vision. The follower records the trajectory points of the leader and tracks the point in this trajectory, which has a predefined safety distance to the leader to ensure cohesion and safe separation of UAVs. Based on the selected point, the motion planning generates a new collision-free reference trajectory for the follower tracking. It is noted that the goal position of the follower varies as its leader moves. The followers will regularly replan their paths online to avoid obstacles in their trajectories. The advantage of this flocking method is to ensure that the followers track their targets and simultaneously achieve obstacle avoidance even when the leader's position is temporarily unknown.

III. Visual Sensing

In order to achieve robust and fast tracking, a markerless vision sensing solution is developed for this flocking system. Although, nowadays, lots of efforts have been made for visual tracking in academia, most of them are still hard to work in real applications due to practical constraints, e.g., limited computational resources and fast varying lighting conditions. As a popular technology in signal processing, a correlation filter has been extended to computer vision societies and also achieved great success in visual tracking applications. Formulated as a discriminative tracker, a correlation tracker achieved very impressive performance in terms of accuracy and efficiency. This paper designs a tracking system on the basis of minimum output sum of squared error (MOSSE) [19] since it provides the highest speed with good performance. Different from MOSSE using a single channel of features, the proposed algorithm employs multilevel-oriented gradient features for robust tracking in thermal and visible images.

A. Discriminative Tracking

A correlation filter models appearance of a target using a filter h trained by a number of grayscale image patches $f_i, i = 1, \ldots, t$ with $M \times N$ pixels centered around the target. The tracker considers all cyclic shift $x_{m,n}, m = 0, \ldots, M-1, n = 0, \ldots, N-1$, as training examples for the classifier [20]. These are labeled with desired Gaussian functions $y_i, i = 1, \ldots, t$, so that $y_i(m, n)$ is the label for $x(m, n)$. A feature map is considered for a signal representation. The single channel correlation filter can then be expressed in a spatial domain by solving a ridge regression problem:

$$E(h) = \frac{1}{2}\sum_{i=1}^{t} \|y_i - h \star f_i\|_2^2 + \frac{\lambda}{2}\|h\|_2^2 \tag{1}$$

where functions f_i, g_i, and h are all of size $M \times N$. $\star$ denotes a circular operator, λ is a regularization parameter ($\lambda \geq 0$), and h is the trained filter for classification.

The filter in (1) provides very limited performance. To improve robustness, discriminative correlation filters can be extended to use multidimensional features. We employ a k-dimensional feature map representation of a signal, such as a two-dimensional (2-D) image, and denote feature dimension number $l \in 1, \ldots, k$ of an image patch f_i by f^l. A multichannel objective function in spatial domain is given by

$$E(h) = \frac{1}{2}\sum_{i=1}^{t} \|y_i - \sum_{l=1}^{k} h^l \star f_i^l\|_2^2 + \frac{\lambda}{2}\sum_{l=1}^{k} \|h^l\|_2^2 \tag{2}$$

where f^l and h^l refer to the lth channel of a vectorized image and filter, respectively. The correlation image y_i is synthetically designed as a desired map with bright peak at the target's center. For translational estimation, we define y_i as a 2-D Gaussian distribution which is centered at the target coordinate $\mathbf{x}_i$ with a radius σ

$$y_i = e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{\sigma^2}}. \tag{3}$$

Solving this multichannel form in the spatial domain is even more intractable. To reduce the problem complexity, we transform the spatial convolution to the element-wise production in frequency domain and solve this equation as

$$H = \frac{\sum_{l=1}^{k} F^l \odot Y^{l*}}{\lambda + \sum_{l=1}^{k} F^{l*} \odot F^l} \tag{4}$$

where capital letters denote discrete Fourier transforms of corresponding functions. For instance, H and Y are the Fourier transforms for filter h and map y, respectively. $\odot$ denotes element-wise production and $*$ is complex conjugate. The tracking is carried out on an image patch z in a new frame with a search window size $M \times N$ by computing the response

$$\gamma = \mathcal{F}^{-1}(F \odot H^*) \tag{5}$$

where γ is a response map and $\mathcal{F}$ is Fourier transform. To accommodate target variation, the tracker should have the capability of online learning. We simplify the numerator and denominator of filter H in (4) as A_t and B_t, respectively. The updating of A_t and B_t at time t is

$$\begin{cases} A_t = (1-\eta)A_{t-1} + \eta \sum_{l=1}^{k} F_{t-1}^{l} \odot Y_{t-1}^{l*} \\ B_t = (1-\eta)B_{t-1} + \eta \sum_{l=1}^{k} F_{t-1}^{l*} \odot F_{t-1}^{l} \end{cases} \tag{6}$$

where η is a learning rate parameter. The final response γ of the filter in the translation level with an observation z can be computed as

$$\gamma_t = \mathcal{F}^{-1}\left\{ \frac{\sum_{l=1}^{k} A_{t-1}^{*} \odot Z_t^l}{B_{t-1} + \lambda} \right\}. \tag{7}$$

For scale estimation, the proposed algorithm formulates scale fitting using a one-dimensional filter. We construct a scale pyramid at the estimated location of translational response. Let $V \times U$ denote the target size in a frame and S is the number of the scale level $n \in \{\lfloor -\frac{S-1}{2} \rfloor, \lfloor -\frac{S-3}{2} \rfloor, \ldots, \lfloor \frac{S-1}{2} \rfloor\}$. The tracker extracts an image patch I_n with size of $a^n V \times a^n U$ centered around the estimated location of the translational tracker. a is a scaling factor. We unify all patches in a pyramid with size $V \times U$ and construct the feature pyramid. The training sample is then set to be a rectangular cuboid of the feature pyramid. The cuboid is of size $V \times U \times S$ and centered at the estimated location of the target. We update the scale filter holistically to capture the scale variation of the target. The scale filter is also online learning, refer to (5). A threshold $\mathcal{T}_s$ for the scale is set and the scale filter updates when the max response γ_s is bigger than $\mathcal{T}_s$.

B. Visual Modules Integration

The proposed correlation filter is adopted for discriminative tracking by distinguishing the target from the background. Similarly, it can be extended for global detection due to its inherent property of classification. On the basis of theoretic foundation above, we adopt a correlation filter for visual detection. Since our system requires fully automatic initialization, the initial detector should be trained offline using manually collected images annotated in classifier formulation. To achieve robust target representation and computational efficiency, we employ a feature coding scheme with five levels of oriented gradients for classifier training [21]. Compared to the initial detector, the redetector is able to adapt to appearance variation of the target by using an online learned classifier. To ensure a correct updating, only images with high confidence will be trusted for the redetector training during flight.

Fig. 3 shows the proposed visual sensing framework, including detection, tracking, and redetection modules. A detector H_d

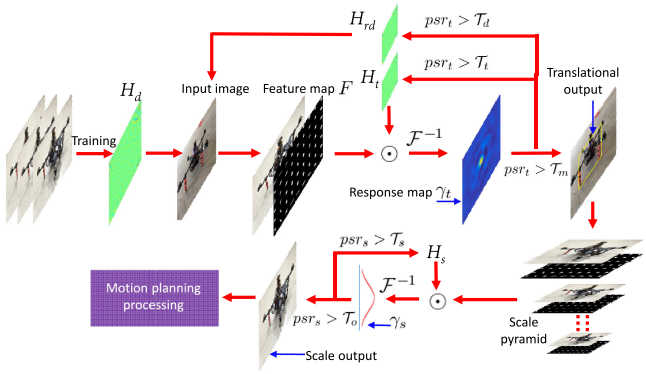


Fig. 3. Sketch map of visual detection and tracking for UAV flocking.

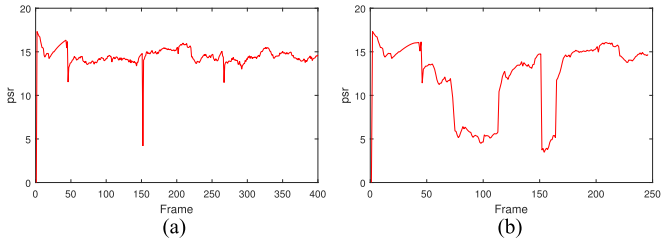


Fig. 4. Confidence response in thermal vision with (a) abrupt motion and (b) occlusion.

will be activated by automatic initialization once a target enters the camera view. As described above, the tracking algorithm consists of translation and scale steps. The initialized target is input to the tracker by extracting multilevel-oriented gradient features (33-levels) [22] for translation estimation first. The peak-to-sidelobe ratio (PSR) is adopted to measure the strength of correlation response. PSR is defined as $\frac{y_{\max} - \mu}{\sigma}$, where $y_{\max}$ is the peak value, and μ and σ are the means and standard deviations of the sidelobe, respectively. It is used as a metric to evaluate the similarity between the target and those candidates.

A translational filter H_t is updated when PSR is larger than a predefined threshold $\mathcal{T}_t$, which is defined for updating redetector H_{rd}. If PSR is larger than another threshold $\mathcal{T}_m$, it means that the target is found by the translational filter. Comparing to H_t, the learning scheme of the redetector is much more conservative because incorrect updating may easily corrupt the filter and result in tracking failure. Therefore, relationship between the thresholds is $\mathcal{T}_d > \mathcal{T}_t > \mathcal{T}_m$. A scale pyramid is built for scale estimation after the translational output. The final output (x_t, y_t, w_t, h_t), consisting of center coordinates, width and height of the target, is transformed into a spatial coordinate for motion planning processing. The detailed procedure is given in the pseudo code in Algorithm 1.

Two tests using thermal vision are conducted to verify the effectiveness of the proposed vision framework. The designed tests include abrupt motion and occlusion, which will activate the redetection module for online reinitialization. The first scenario presents a UAV flying in a forest with an abrupt motion which is caused by shutter effect of the thermal camera. The abrupt motion results in tracking failure at Frame #152 (see Fig. 5). The confidence PSR reflects tracking quality. In the test, it drops to a low level at Frame #152 and the target is

Algorithm 1: Framework of Visual Sensing.

Visual Module ← Image Streaming Input
Initialization
$\mathbf{x}_0 \leftarrow$ target state (x_0, y_0, w_0, h_0) by detector H_d
$\mathbf{Flag}_{H_{rd}} = 0$, local patch $\mathcal{A}(\mathbf{x}_t)$, Feature map F
for $t = 1 : N$ **do**
  if $\mathbf{Flag}_{H_{rd}} = 1$ **then**
    $\mathbf{x}_t \leftarrow H_{rd}$, $\mathbf{Flag}_{H_{rd}} \leftarrow 0$, $F(\mathcal{A}(\mathbf{x}_t))$
  else
    Extracting $F(\mathcal{A}(\mathbf{x}_{t-1}))$
  end if
  Translational Estimation
  if $t = 1$ or $\mathbf{Flag}_{H_{rd}} = 1$ **then**
    H_t, $H_s \leftarrow 4$ for filters training
  else
    $\gamma_t \leftarrow 7$ for response map generation
    Computing confidence psr_t
  end if
  if $psr_t > \mathcal{T}_m$ **then**
    Output Translation results in $(x_t, y_t, w_{t-1}, h_{t-1})$
    Go to Scale Estimation
  else
    Tracking failure, and $\mathbf{Flag}_{H_{rd}} = 1$
    Continue
  end if
  Updating filter $H_t \leftarrow 6$ if $psr_t > \mathcal{T}_t$
  Updating Re-detect filter $H_{rd} \leftarrow 6$ if $psr_t > \mathcal{T}_d$
  Scale Estimation
  Formulating scale pyramid and estimating scale response γ_s
  Updating filter H_s if $psr_s > \mathcal{T}_s$
  Output scale results if $psr_s > \mathcal{T}_o$
  Output Tracking result $\mathbf{x}_t$ in (x_t, y_t, w_t, h_t)
end for

lost. The redetector is then triggered [see Fig. 4(a)] for target detection and PSR returns to a high level after the target is recaptured successfully. Second, we present test scenarios involving two occlusions that occurs at around Frame #112 and #153. As shown in Fig. 4(b), PSR of the tracker declines and maintains a low level during the occlusion but it can return to a normal value once the target is redetected after occlusion. Due to page limitations, we only show image samples at the first occlusion in Fig. 5. It is observed that PSR can quickly return to a stable level after a successful redetection. That is because the online trained filter of the tracker is retained after the target loses. It records the target information during tracking process, and double check the correctness of results generated by the redetector before the retracking.

C. Waypoint Generation

The tracking results will output the target status including its center coordinate and size in the image. Using these parameters and the known physical size of the target, we can compute its three-dimensional (3-D) location in the camera coordinate with distance and attitude angle, given the camera intrinsic parameters. We need to transform the image coordinate with pixel unit to the camera plane. As shown in Fig. 6(a), the target has physical size L and image with size of l. According to the triangle property, we have the relationship of $\frac{d}{L} = \frac{f}{l}$, where d is the distance between the camera and the target, and f indicates the focal length. It can be observed that the distance d is inversely proportional to the size l in image plane. Based on this relation, d is calculable since L and f are given and l is measured in an image plane. For attitude angle, it includes angle in elevation and azimuth direction in spatial. As shown in a 2-D geometry map in Fig. 6(a), attitude angle is $\theta = \arctan \frac{\overline{qc}}{f}$.

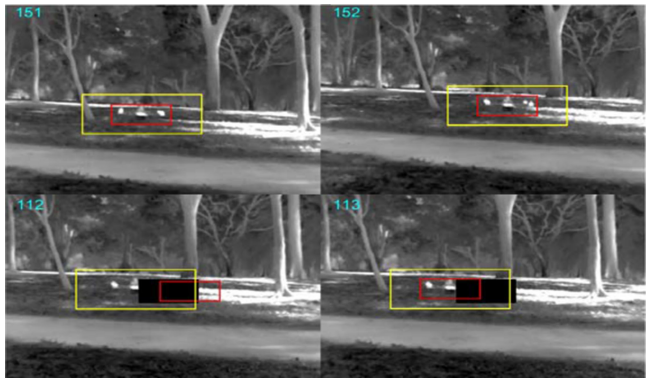


Fig. 5. Image samples in thermal vision with abrupt motion (first row) and occlusion (second row). Tracking result is marked in red and detection result is given in yellow.

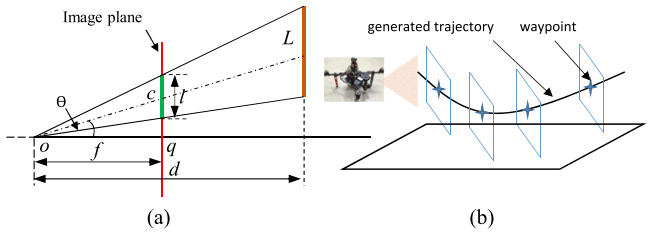


Fig. 6. (a) Image geometry for distance estimation. (b) Waypoint generation.

With the measured distance and attitude of the tracked target, its spatial coordinate can be computed. As shown in Fig. 6(b), tracking results transform to the waypoints for trajectory generation. It should be noted that the system adopts discriminative-based tracker for tracking due to its robust adaption of target's appearance changing. Because of the relative motion of the target to the camera, it is hard to keep a consistent view of the target, which results in target's appearance variation. Therefore, it is difficult to estimate an accurate pose of the target in spatial. However, our tracker provides a robust target sensing which can guide the follower flying to the target. An approximate estimation of waypoints is acceptable for our autonomous flocking system.

IV. Navigation System

A. Localization and Mapping

To realize autonomous flight in GPS-denied environments, a UAV should be able to sense the surrounding environment using

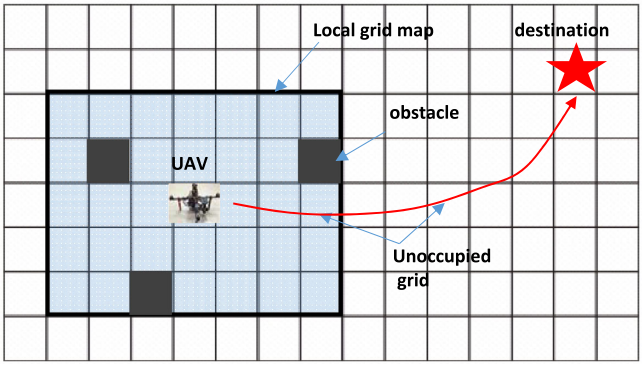


Fig. 7. Local occupancy grid map and global map.

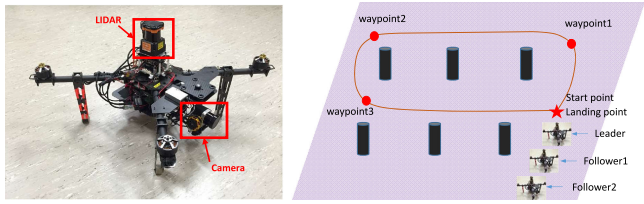
on-board sensors. In this flocking system, we install a LiDAR for each individual UAV to sense the surrounding information for localization, map generation, and obstacle avoidance. Since our flocking system is designed to operate in cluttered environments, like urban and forest, a feature-based 2-D SLAM algorithm is implemented [23]. The basic ideas can be summarized as: 1) filter out long straight lines, which may be ground or wall in urban environments; 2) extract ellipses of tree trunks in forest and estimate their position and size in every scanning; 3) match scans based on a nearest neighbor algorithm, also the size of ellipses will be considered; 4) estimate UAV pose based on the scan to scan transformation; 5) build a pose graph based on matching results and use the sliding window technology to formulate a local graph instead of a global one due to speed consideration; and 6) use the G2O library to optimize the pose graph. It is worth noting that the LiDAR is fixed on UAVs without a stabilizer. Our feature-based SLAM algorithm can handle the small body slope.

The point clouds provided by the LiDAR are employed to generate a 3-D occupancy grid map centered in the UAV, which is divided evenly into cubes to represent the environment. Each cube can be labeled in two statuses: occupied or unoccupied. The occupied cubes and their surrounding area will be blocked for flight in order to keep a safe motion space. Only the rest unoccupied area can be used for path planning.

The explored environment is assumed unknown and the computational load of the motion planning will increase with the size of the map. For real-time operation, a local grid map is developed on the basis of measured data as shown in Fig. 7. A sensing area with fixed size centered at the UAV is represented by the occupancy grid map. The rest of area outside the local map is represented by unoccupied spaces. The local zone moves with the UAV and its size depends on the sensing range of LiDAR. The resolution of the occupancy grid map is a tradeoff between the accuracy of environment representation, and computational resources of the UAV.

B. Motion Planning

Motion planning is important for autonomous systems. We adopt an online motion planning method [18] for UAV exploration, which is able to generate a collision-free path and trajectory in real time. The online path planning employs an improved



(a) (b)

Fig. 8. (a) Platform of quadrotor UAV. (b) Flight scenario of flocking.

A* searching algorithm to find a collision-free path from a start point to a destination point on the local grid map. Since the A* algorithm works on the local grid map which may change during the movement of the UAV, the generated path may not achieve the global minimum, but its efficiency can be improved significantly.

The path planning module does not consider the kinematic model of the UAV. Therefore, online trajectory generation is necessary to bridge the high-level path planning with the low-level motion control of the UAV. We employ Reflexxes Type IV Motion Library [24] to generate 3-D trajectories for UAVs. The trajectory generator is able to compute an entire time-optimal trajectory from an arbitrary initial state of motion to a given target subject to velocity and acceleration limitations within 1 ms. As the Reflexxes trajectory generator does not consider obstacle avoidance and it is open loop, the generated trajectory has a chance to cause collision. It is necessary to check collision of the trajectory. A local target generator is adopted to perform trajectory collision checking, which is an intermediate element between the path planning and the trajectory generator. The online path planning and trajectory generation run in 10 and 50 Hz, respectively, since the online path planning requires more computational resource than that of the trajectory generator.

V. Experiments

To verify the effectiveness of the proposed flocking system, flight experiments have been conducted by using in-house-built quadrotor UAVs, which can fly about 15 min with 1.5 Kg extra payload by using an 8000 mAh LiPo battery. It is equipped with an LiDAR sensor: Hokuyo UTM-30LX [see Fig. 8(a)], which can provide 0.1–30 m measurement range, 270° sensing angle, and 40 Hz scanning rate. The drone is able to use two types of vision sensors: an Optirs PI450 thermal camera and a PointGrey BFLY-PGE-12A2C-CS day camera.

The UAV has a customized flight control computer and an AscTec Mastermind mission computer. The flight control computer has been used to perform low-level feedback control and trajectory tracking. The mission computer is employed to integrate the visual sensing system with the navigation system to realize high-level command and control, such as the flocking mission, based on the structure illustrated in Fig. 2. Software integration has been implemented by using a middleware: robot operating system (ROS) due to its good modularity and concurrent resource handling. The aforementioned algorithms have been implemented in individual ROS nodes, and standardized messages have been used among those nodes.

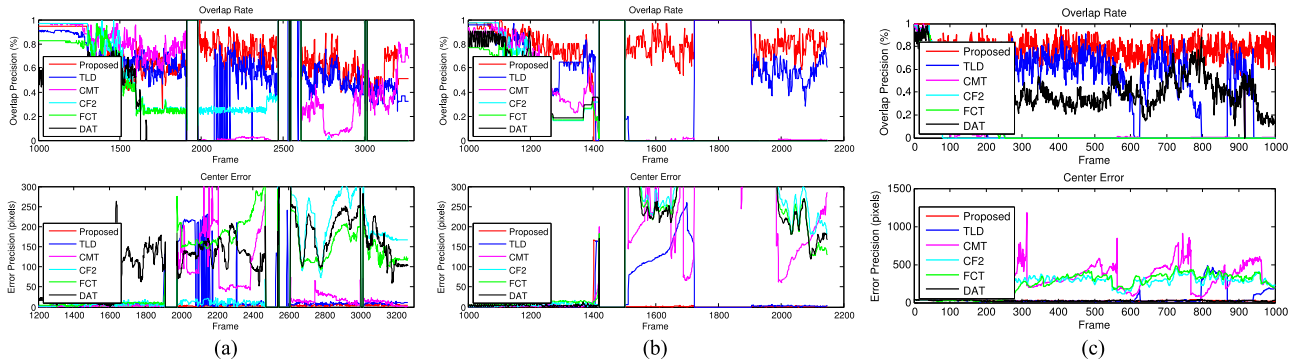


Fig. 9. Precision performance for the (a) *Indoor1* (thermal) and (b) *Indoor2* (visible). (c) Precision experiment for *Outdoor*. First row shows the overlap rate precision and the second row presents the center error precision.

The waypoints are predefined in the initial stage and only the team leader knows its task information. On the other hand, a follower tracks its leader and dynamically plans its path for the tandem flocking. Initially, UAVs take off 3 m apart and with 1.5 m height. Their distance intervals range from 3 to 8 m during flight. The images captured by the on-board cameras are normalized with resolution of 382×288. The cameras are calibrated in advance for target state estimation and also stabilized by using gimbals.

A. Visual Tracking

In this system, visual sensing is used to measure the motion status of the leader UAV and aid the followers to track its corresponding leader's steps. We apply our algorithm on thermal and visible cameras, respectively. The communication model will be activated only in case when the visual system fails to detect the target status. Two rounds of indoor real flight with thermal and day cameras are applied first, respectively, to verify the performance of proposed vision system. Selected state-of-art algorithms including TLD [12], CMT [25], CF2 [26], FCT [27], and DAT [28], are applied for comparison.

The thermal experiment is conducted first and the flocking scenario is shown in Fig. 8(b). The UAV team operates at an indoor scene in size of 35×35 m. All the UAVs are well initialized and followers detect the target automatically. The total number of the frames is 3300 and the UAVs take off around the Frame #1200. Tracking accuracy is evaluated using center error and overlap ratio [29]. This experiment is shortened in *Indoor1*.

At the early stage of the experiment, the target stands on the ground for a while before it takes off. All algorithms perform well except DAT, which shrinks and drops tracking accuracy significantly even when the target is static in an image. After taken off, the flocking team flies straightly in 10 m to Waypoint #1, followed with a left turn. DAT totally loses target at Frame #1627. This is because DAT relies on a color feature which is unstable in thermal images. Thermal camera provides a heat distribution image of the environment and the object is recognizable based on its contour. CF2 and FCT cannot present good performance in overlap rate after the UAV takes off. It is understandable for CF2 because it does not consider scale change. Although FCT tracks the scale of the target, it cannot give a robust scale estimation when the target is far away from the camera. As shown in Fig. 9(a), the proposed tracker, TLD, and CMT are stable before the first turning corner at Frame #1910.

The leader turns left first at Waypoint #1, while followers follow and enter a curve. For a fair evaluation, we set the overlap rate and center error to optimal values when the target disappears in this moment. Redetection of the proposed algorithm is activated for target searching, and it recaptures the target at Frame #1977. TLD, however, loses the target after two frames tracking and it redetects the target at Frame #2080. It performs very unstable from Frame #2080 to Frame #2600 and it repeats the processing of tracking, losing, and redetection in this period. Occasionally, TLD presents false positives because of the attraction of outliers [see Fig. 10(a)]. The failure of TLD may be caused by poor resolution of thermal images. In contrast, the proposed algorithm performs very stable and robust because of using contour-based multilevel-oriented gradient features and the correlation filter. CMT recaptures the target at Frame #2622, but cannot provide a good tracking, because keypoint-based global matching may not be stable in thermal image. CF2 and FCT lose target at the first turning corner. Even CF2 retracks target by chance but it does not perform well in the rest of the sequences.

A day camera has been employed in the second round under the same scenario (*Indoor2*). As shown in Fig. 9(b), the UAV team takes off at Frame #1090 and all the trackers keep tracking before the leader disappears at Frame #1416. The proposed algorithm can promptly recapture the target when the target reappears in image and it is stable through the experiment. In contrast, the other trackers cannot recover even the target is reappearing. Among of them, TLD retracks the target at Frame #1905, thanks to its detection module.

Additionally, we present a challenging outdoor experiment with a day camera to further verify the effectiveness of the proposed tracker, which is abbreviated as *Outdoor* conducted in a small forest with size of 50×50 m. In total, 5000 frames have been collected with a resolution of 644×482. Because of time limit, we only label frames with $\#5\times$ and the final labeled images are 1000. The experiment shows a scenario that the UAVs fly in a forest. Since some trackers are designed for local tracking, this experiment does not consider occlusion and target disappearing even it is a false assumption in the practical application. All the trackers are well-initialized. After taking off, CF2 loses target at Frame #405 because it is attracted by

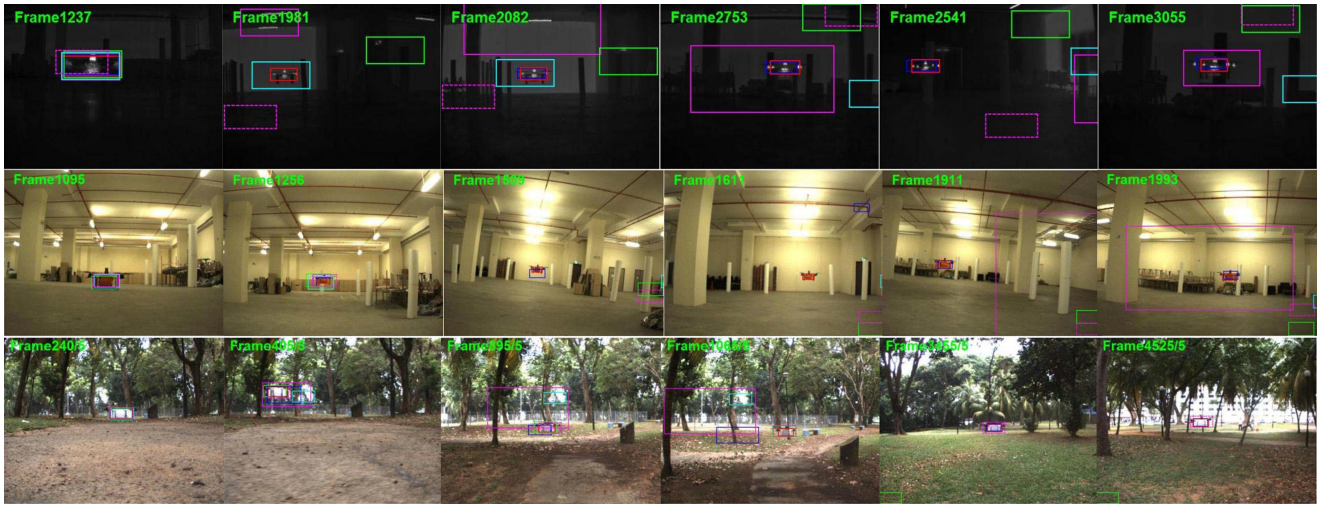


Fig. 10. Representative image results of tracking for *Indoor1*, (first row), *Indoor2* (second row), and *Outdoor* (third row). The proposed algorithm, TLD, CMT, CF2, and FCT are in red, blue, magenta, cyan, and green, respectively. DAT is in magenta with dash line.

TABLE I
QUANTITATIVE RESULTS IN MEAN ERROR

Mean Center Error						
Video	DAT	FCT	CF2	CMT	TLD	Prop
Indoor1	75.14	61.89	46.35	23.41	14.96	**6.588**
Indoor2	129.3	128.9	138.6	123.4	15.87	**7.146**
Oudoor	**23.55**	236.3	243.9	287.2	64.12	24.01
Average	76.01	142.3	142.96	144.7	31.65	**12.58**
Mean Overlap Rate						
Indoor1	0.3426	0.4932	0.5803	0.6538	0.6990	**0.7843**
Indoor2	0.5385	0.4975	0.5997	0.5860	0.7745	**0.8895**
Outdoor	0.4325	0.1376	0.0610	0.0631	0.4836	**0.7831**
Average	0.4379	0.3761	0.4137	0.4343	0.6524	**0.8190**

the background with similar appearance to the target. CMT and TLD also drift after Frame #428. FCT loses the target at Frame #1167 because of the cluttered background. TLD retracks the target at Frame #1260 using the redetection function. DAT keeps tracking the target through the entire sequence. This experiment is difficult because of very complicated background and various illuminations in the forest. The proposed algorithm, however, keeps a successful tracking with high overlap ratio because the powerful integration of multilevel gradient features and the correlation filter.

We also present a mean error in quantitative results as shown in Table I. The best and second results are reported in bold and underline. It can be observed that the proposed tracker achieves the best performance and significantly outperforms the other algorithms in terms of center error and overlap rate. The TLD obtains the second place in experiment. It is noted that the DAT has the best score at *Outdoor* in center error. However, the DAT is only third in performance in overlap rate at *Outdoor*. In contrast, the proposed tracker has the best and the second score in overlap rate and center error, respectively, and it outperforms DAT significantly in terms of overlap rate at *Outdoor*.

B. Environment Exploring

Different logics have been applied to the leader and the followers for navigation and obstacle avoidance. The leader will follow the predefined waypoints and avoid obstacles along the online generated trajectory simultaneously. Waypoints for the followers will be generated dynamically in terms of the measured target position, which may not be stable for the online trajectory generation sometime. Therefore, we propose an alternative solution: 1) store all the positions of the leader with time stamps in memory; 2) find a suitable point from these positions, and make sure the distance between this point and the leader is 3 m at least. If such a point does not exist, choose second safe point based on leader's moving direction; and 3) monitor the position of the leader from image measurement and SLAM estimation via communication. If camera tracking is failed, we switch to the SLAM estimation, but tracking error may drift over time.

A cost occupancy grid map will be created from laser scan data, and then a path will be computed by using the A* algorithm. To meet the dynamic motion constraints of the UAVs, the reflexxes library is used to generate a time optimal and smooth trajectory based on the path. Every 0.02 s, the generated trajectory reference will be sent to the flight controller to realize the closed-loop control.

Two flight tests have been conducted with visible and thermal cameras, respectively. The flight trajectories are illustrated in Fig. 8(b). Three UAVs are used for visible experiment. UAV #1 is the leader, and UAV #2 and #3 are the followers, with origin points (7, 0), (3.5, 0), and (0, 0). Two UAVs are used for the thermal experiment with leader UAV #1 and follower UAV #T2, and their starting points are (7, 0) and (3.5, 0).

Fig. 11(a) shows the performance of UAV #1. As it just followed the preplanned waypoints and did not consider the other two UAVs' positions, its trajectory fits the straight lines between the waypoints perfectly. From the quality of map alignment, the SLAM algorithm for UAV #1 also performs great. While for the result of UAV #2 and #3 in Fig. 11(b) and (c), the SLAM has drifted. That is mainly because the other moving UAVs interfered with the feature-based scan matching. To make sure the safety of whole system in the case of vision tracking failure or having a large error, path planning compares the position

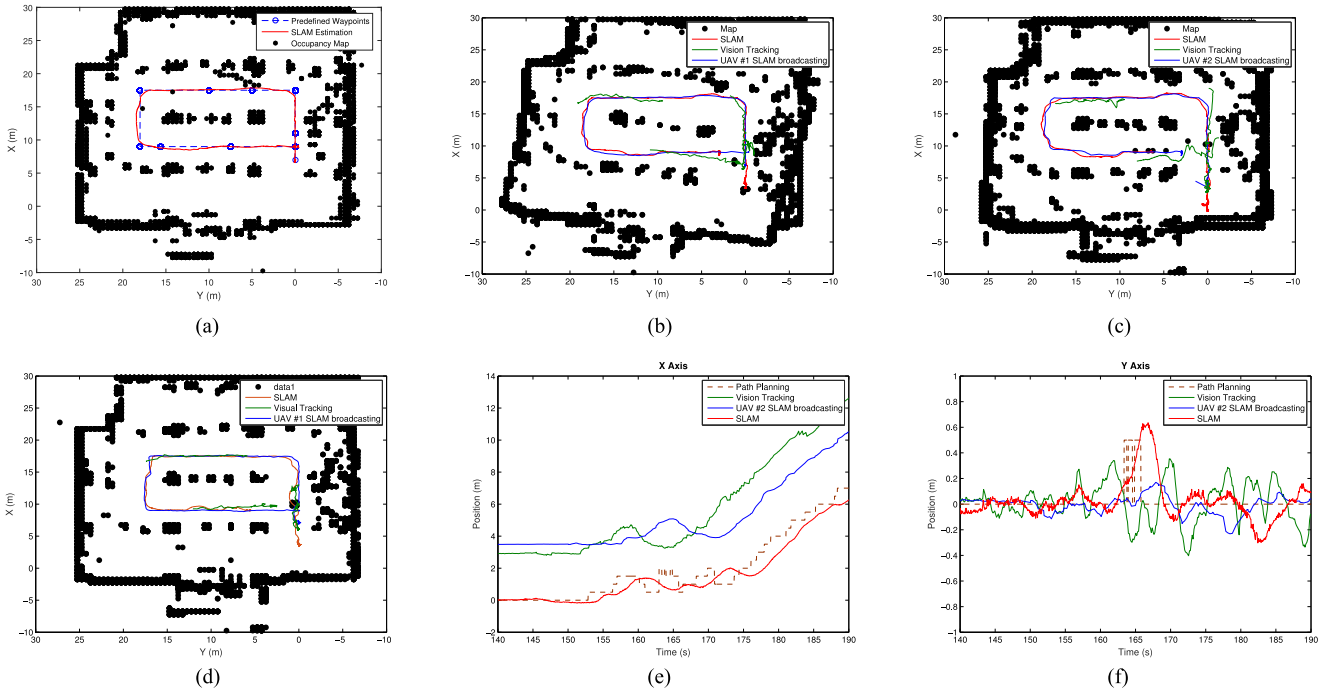


Fig. 11. (a) Trajectory of UAV #1 in test 1. (b) Trajectory of UAV #2 with a visible camera in test 1. (c) Trajectory of UAV #3 with a visible camera in test 1. (d) Trajectory of UAV #T2 with a thermal camera in test 2. (e) *x*-axis position of UAV #3. (f) *y*-axis position of UAV #3.

difference between the vision tracking result and the leader's broadcasting SLAM data. If the difference is too large, in our case, the max tolerance error is 1.41 m, the algorithm will trust the latter one. In fact, vision tracking displays very stable performance if the target presents the images. While the target was turning around a pillar, it might be blocked. The vision tracking would not get any valid information and then fail. From Fig. 11(b), (c), and (d), we can see that the vision tracking trajectories are discontinued, because the target is hidden behind obstacles or out of field of views of the cameras, when UAVs are turning around the obstacles. For this case, our algorithm will switch to leader's SLAM guidance.

The capability of collision avoidance between moving UAVs is shown in Fig. 11(e) and (f), where SLAM line can be considered as the measured position of the UAV and the lines of vision tracking and UAV #2 broadcasting are both target reference positions. In other words, UAV #3 has two channels information of its target UAV #3 from measure of the camera and broadcasting. Path planning line shows the final reference trajectory for UAV #3, which is sent to the outer-loop of the flight control system. In this experiment, UAV #3 tries to step back to keep a 3 m distance from UAV #2 during 160–175 s. All the UAVs take off at about 145 s, and UAV #2 went to (5, 0) first around 159 s and then fell back to (3.5, 0) around 165 s. Note that the broadcast position of UAV #2 has a delay because of the heavy communication load. UAV #3 detects this issue by the vision tracking first. In this case, the path planning algorithm trusted the vision tracking result instead of UAV #2 broadcasting, because their difference is small. At around 167 s, the path planning algorithm chose a safe position in both in *x*- and *y*-axis [from (2, 0.5) to (0.5, 0)] to avoid collision. Based on this experiment, it can be proven that visual tracking is effective and reliable to maintain safe flight for UAV flocking. Another issue needed to note is that the maps in Fig. 11(a) to (d) are generated online, which means dynamic obstacles and outliers are also shown here. So it is reasonable that the paths pass through the obstacles in the map.

Those real flight experiments have successfully verified the effectiveness of the developed flocking system including sophisticated algorithms, such as vision tracking, SLAM, and path planning. All of the algorithms are processed in an on-board computer with i7-3612QE (4 × 2.1 GHz) CPU and 4G RAM running in Ubuntu 14.04. To balance the performance of those computational intensive algorithms, the frequency of laser scan is limited to 20 Hz, so that all the algorithms can run in full speed (SLAM, 3-D mapping, path planning, and visual tracker are running at 20, 20, 10, and 25 Hz, respectively) without overload.

VI. Conclusion

In this paper, we presented a sophisticated flocking system, which successfully integrated various advanced technologies, including LiDAR-based SLAM, and a visual system for sensing in both day and night without continuous wireless communication and GPS signals that are required in traditional flocking systems. To address the enhanced visual sensing, a multichannel correlation filter based visual algorithm was developed for the robust tracking in thermal and visible visions. Enhanced tracking-learning-detection fashion was proposed for long-term robust visual tracking. Environment understanding and path planning modules were implemented for autonomous task operation. The experiments at indoor and outdoor scenarios

verified the effectiveness of the proposed visual algorithm for challenging practical applications.

The developed flocking system realized a prototype of multi-UAV autonomous flight. For practical applications in future, we are going to further optimize the system performance, including increasing the flight speed and enhancing SLAM capabilities by using 3-D sensing, etc.

REFERENCES

[1] O. O'Loan and M. Evans, "Alternating steady state in one-dimensional flocking," *J. Phys. A: Math. Gen.*, vol. 32, no. 8, pp. L99–L105, 1999.

[2] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," *ACM SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 25–34, 1987.

[3] R. O. Saberand R. M. Murray,"Flocking with obstacle avoidance: cooperation with limited communication in mobile networks," in *Proc. 42nd IEEE Conf. Decision Control*, 2003, vol. 2, pp. 2022–2028.

[4] H. G. Tanner, "Flocking with obstacle avoidance in switching networks of interconnected vehicles," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, vol. 3, pp. 3006–3011.

[5] H. Shi, L. Wang, and T. Chu, "Virtual leader approach to coordinated control of multiple mobile agents with asymmetric interactions," *Physica D*, vol. 213, no. 1, pp. 51–65, 2006.

[6] W. Burgard, M. Moors, C. Stachniss, and F. E. Schneider, "Coordinated multi-robot exploration," *IEEE Trans. Robot.*, vol. 21, no. 3, pp. 376–386, Jun. 2005.

[7] W. Dong, "Flocking of multiple mobile robots based on backstepping," *IEEE Trans. Syst., Man., Cybern. B*, vol. 41, no. 2, pp. 414–424, Apr. 2011.

[8] T. S. No, Y. Kim, M.-J. Tahk, and G.-E. Jeon, "Cascade-type guidance law design for multiple-UAV formation keeping," *Aerosp. Sci. Technol.*, vol. 15, no. 6, pp. 431–439, 2011.

[9] S. Zhao *et al.*, "A robust real-time vision system for autonomous cargo transfer by an unmanned helicopter," *IEEE Trans. Ind. Electron.*, vol. 62, no. 2, pp. 1210–1219, Feb. 2015.

[10] M. Mueller, G. Sharma, N. Smith, and B. Ghanem, "Persistent aerial tracking system for uavs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 1562–1569.

[11] J. Pestana, J. L. Sanchez-Lopez, P. Campoy, and S. Saripalli, "Vision based GPS-denied object tracking and following for unmanned aerial vehicles," in *Proc. IEEE Int. Symp. Safety, Security, Rescue Robot.*, 2013, pp. 1–6.

[12] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[13] C. Fu, A. Carrio, M. A. Olivares-Mendez, R. Suarez-Fernandez, and P. Campoy, "Robust real-time vision-based aircraft tracking from unmanned aerial vehicles," in *Proc. IEEE Int. Conf.Robot. Autom.*, 2014, pp. 5441–5446.

[14] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 983–990.

[15] A. Nussberger, H. Grabner, and L. Van Gool, "Robust aerial object tracking in images with lens flare," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 6380–6387.

[16] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared LEDs," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 907–913.

[17] S.-M. Hung and S. N. Givigi, "A Q-learning approach to flocking with UAVs in a stochastic environment," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 186–197, Jan. 2017.

[18] F. Liao *et al.*, "3D motion planning for uavs in gps-denied unknown forest environment," in *Proc. IEEE Intell. Veh. Sympo.*, 2016, pp. 246–251.

[19] D. S. Bolme *et al.*, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 2544–2550.

[20] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 1090–1097.

[21] H. K. Galoogahi, T. Sim, and S. Lucey, "Multi-channel correlation filters," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 3072–3079.

[22] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[23] J. Q. Cui, S. Lai, X. Dong, and B. M. Chen, "Autonomous navigation of UAV in foliage environment," *J. Intell. Robot. Syst.*, vol. 84, nos. 1–4, pp. 259–276, 2016.

[24] T. Kröger, *On-Line Trajectory Generation in Robotic Systems: Basic Concepts for Instantaneous Reactions to Unforeseen (Sensor) Events* volume 58 of Springer Tracts in Advanced Robotics. New York, NY, USA: Springer, 2010.

[25] G. Nebehay and R. Pflugfelder, "Clustering of static-adaptive correspondences for deformable object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 2784–2791.

[26] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3074–3082.

[27] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2002–2015, Oct. 2014.

[28] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition, 2015*, pp. 2113–2120.

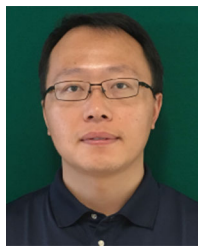
[29] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

Yazhe Tang received the Ph.D. degree in robotics from the Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Hong Kong, in 2013.

Before joining City University of Hong Kong, he was a Research Assistant with the Korea Atomic Energy Research Institute, Daejeon, South Korea, from September 2007 to August 2009. He is currently a Research Scientist with Temasek Laboratories of National University of Singapore, Singapore. He has also been an Assistant Professor with the Department of Precision Mechanical Engineering, Shanghai University, Shanghai, China, since 2014. His research interests include robot sensing, visual tracking, visual saliency, and omnidirectional vision.

Yuchao Hu received the B.E. degree in automation and the M.E. degree in pattern recognition and intelligent systems from Northeastern University, Shenyang, China, in 2012 and 2014, respectively.

Since December 2014, he has been an Associate Scientist with Temasek Laboratories, National University of Singapore, Singapore. His research interests include unmanned systems, SLAM, and robot simulation systems.



Jinqiang Cui received the B.S. and M.S. degrees in mechatronic engineering from Northwestern Polytechnical University, Xi'an, China, in 2005 and 2008, respectively. He received the Ph.D. degree in electrical and computer engineering from National University of Singapore (NUS), Singapore, in 2015.

His Ph.D. research focuses on navigation of unmanned aerial vehicles (UAV) in GPS-denied environments, especially forest. He was a Research Scientist with the Control Science Group, Temasek Laboratories, NUS, from 2015 to 2016. He is now serving as the CTO of a startup company specializing in the development of UAV navigation systems.

Fang Liao received the B.E. and M.E. degrees in control and navigation from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1992 and 1995, respectively, and the Ph.D. degree in control and navigation from the Nanyang Technological University, Singapore, in 2003.

During 1995 to 1999, she was an Engineer with the Research Institute of Unmanned Air Vehicles, Beijing University of Aeronautics and Astronautics. From 2002 to 2004, she was a Research Associate and then a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University. She has been with Temasek Laboratories, National University of Singapore, Singapore, since 2004 and currently is a Senior Research Scientist. Her research interests include motion planning, cooperative control, multiagent systems, robust and adaptive control theories and application, fault tolerant control, and constrained optimization methods.

Mingjie Lao received the Bachelor's degree in mechanical engineering from National University of Singapore in 2015.

From May 2012 to August 2012, he did an internship with the MIT-Singapore Alliance of Research and Technology to develop a virtual reality system for remote control of autonomous vehicle. He joined Temasek Laboratories, National University of Singapore in July 2015. His main research interests include robot vision, vision tracking, and robotics control.

Feng Lin received the B.Eng. degree in computer science and control and the M.Eng. degree in system engineering from Beihang University, Beijing, China, in 2000 and 2003, respectively. He received the Ph.D. degree in computer and electrical engineering from the National University of Singapore (NUS), Singapore, in 2011.

He is currently a Senior Research Scientist with the Temasek Laboratories, NUS, and a Research Assistant Professor with the Department of Electrical and Computer Engineering, NUS. His main research interests include unmanned aerial vehicles, vision-aided control and navigation, target tracking, robot vision as well as embedded vision systems.

Dr. Lin was the recipient of the Best Application Paper Award, 8th World Congress on Intelligent Control and Automation, Jinan, China (2010).

Rodney S. H. Teo received the B.Eng. degree in mechanical engineering from the National University of Singapore, Singapore, in 1990, and the M.S. and the Ph.D. degrees in aeronautics engineering from Stanford University, Stanford, CA, USA, in 1998 and 2004, respectively.

He has held positions as a Project Engineer (1990–1995) and a Project Manager (1996–1997) on helicopter acquisition and system integration projects in the Defence Materiel Organisation of Singapore. He is currently a Senior Research Scientist of the Temasek Laboratories, National University of Singapore. His current research interests include research and development in the area of autonomy for miniunmanned aerial systems.