Visual Navigation for UAVs Landing on Accessory Building Floor

Jyi-Shane Liu[1, 2]
[1]Department of Computer Science
National Chengchi University
Taipei, Tawian
[2]Pervasive Artificial Intelligence Research Labs, Taiwan
liujs@nccu.edu.tw

Hsiao-Che Liu
Department of Computer Science
National Chengchi University
Taipei, Tawian
107753028@nccu.edu.tw

***Abstract*—Landing is a critical step in most real world UAV applications, especially in delivery. A package delivery is successful only when a landing or a low-altitude drop-off (pseudo-landing) is completed. For precision landing requirement, vision-based navigation techniques are of high potential to be reliable and accurate. In this paper, we present a research work on autonomous visual navigation for precision landing on accessory building floor. We incorporate some state-of-the-art vision-based methods, develop other functional components to present an employable autonomous navigation system for precision landing near buildings. Initial experiments in a real world scenario show an encouraging results with high success rate of performing precision landing.**

Keywords-drone navigation; vision-based navigation; behavior tree; precision landing; building accessory floor

I. INTRODUCTION

Landing is a critical step in most real world UAV applications, especially in delivery. A package delivery is successful only when a landing or a low-altitude drop-off (pseudo-landing) is completed. For any UAV operation, landing is the last step of an executive process of taking-off to a certain altitude, traversing an aerial pathway, and coming back to ground for (temporary) halt. Unsuccessful landing presents a risk of package loss or equipment damage, both a great loss of task value and business property. The central requirement for a successful multi-rotor typed UAV landing is location precision. A reliable and accurate location identification method must be provided in order for the autonomous drone to pinpoint the exact location of a designated landing site. Furthermore, this location identification method also needs to be able to estimate the difference between current location and target location so that a movement plan to reduce and eliminate the location difference can be calculated.

Considering the UAV as a machine with high mobility in space, location identification method is, in fact, indispensable for any UAV operations, from home base localization, flight path traversal, to landing spot lock-in. In this regard, GPS sensor and GPS signal have been commonly used in all types of UAVs. As a mostly reliable global localization system, GPS provides universal location information in a world wide geographic coordinate system. Therefore, GPS-based navigation has been the dominant approach adopted in most UAV applications [1]. While mostly successful in open fields and remote areas, GPS-based navigation encounters a number of environmental and technical restrictions. First, GPS signals may be lost or interfered. Urban environments with crowded buildings, man-made structures and intensive electric facility present serious threat to reliability of GPS-based navigation [2]. Second, commercial GPS devices, especially consumer-grade products, do not provide high-level positioning accuracy. It was reported that the horizontal accuracy for civilian GPS is 10~15 meters at 95% confidence interval and the vertical accuracy is even worse [3]. As a result, GPS-based navigation often suffers from deviation/fluctuation from the desired path and becomes impractical for landing operation that is critically dependent on horizontal and vertical position precision.

In recent years, there has been increasing research interest in vision-based navigation for UAVs landing [4][5]. Vision provides universal perception and induces rich information on surrounding environments. For precision landing requirement, vision-based navigation techniques are also of high potential to be reliable and accurate. Landing location can be specified by a spot marker with unique graphic pattern [6] or colored area [7]. Visual object identification and tracking techniques can then be used to recognize the landing spot and effectively navigate the drone toward precision landing. Previous work includes different technical aspects of a successful landing phase operation such as 3D terrain map generation for locating secure landing area [8], accurate marker detection and tracking [9], and accurate landing with ground pattern recognition and adaptive landing strategy [10]. Precision landing has also been extended to high-risk landing spots such as ship board [11] and moving platform [2]. Indeed, there are a wide variety of landing spot conditions and much research work still needs to be done.

In this paper, we present a research work on autonomous visual navigation for precision landing on accessory building floor. Our technical contribution is in two folds. First, we address the problem from a system perspective and design a complete technical solution. We incorporate some state-of-the-art vision-based methods, develop other functional components to present an employable autonomous navigation system for precision landing near buildings. Second, we implement the technical design, use behavior tree as control architecture to integrate all functional components, and employ the developed system on a quadrotor for field test. Initial experiments in a real world scenario show an encouraging results with high success rate of performing precision landing.

978-1-6654-0483-9/20/$31.00 ©2020 IEEE
DOI 10.1109/ICPAI51961.2020.00037

II. Landing On Accessory Floor Area

For package delivery service, a UAV typically flies over a long distance with multiple waypoints navigation to reach the final destination and lands the package on the ground at the designated spot. Precision landing is critical to package reception and to fulfill the task value of package delivery. In urban environments, package reception spot is mostly favored at flat area around buildings for convenient human access and package retrieval. Due to potential signal interference and insufficient position precision, GPS-based navigation for landing area surrounded by man-made structures becomes unreliable and failure prone. An accessary building floor is typically surrounded by building structures in some directions, thereby posing challenges to landing with precision and safety. We consider visual navigation approaches to address the landing problem near buildings. Assuming that the UAV has been navigated by GPS waypoints to reach the airspace in the proximity of the final destination, the landing process is divided into four steps: (1) recognize the landmark building, (2) orient the UAV to establish a glide path, (3) activate the approach phase, and (4) descent for touchdown. At each step, the UAV's monocular camera image is used as the primary source of navigation information, with the embedded barometer and IMU (Inertia Measurement Unit) as the secondary sensory data sources.

A. Landmark Recognition

Assuming that aerial pictures of the target building are given, landmark recognition is to perform visual check on the target building and ensure the UAV's arrival to the landing site. A set of pictures of the target building at different altitudes, angles, and distances are prepared beforehand. When the final GPS waypoint marking the nearby position of the target building has been reached, the UAV will enter the landing process and activate image recognition for target building as landmark. Once the building's image recognition is successful, the UAV will initiate the second step of the landing process and adjust its position for proper approach. If the image recognition fails, the UAV could either circle around and change altitude for more attempts of image recognition or abort the mission and return to base.

B. Orientation for Glide Path

Landmark recognition on target building is conditionally accepted if one of several pictures in the building's image profile is matched. The hovering position of the UAV at the instance of accepting the landmark recognition may not be the best starting point to activate the approach phase. In order to ensure a safe pathway to approach the target building and arrive at the vertical zone on top of the landing spot, a virtual glide path is pre-planned for the target building. This virtual glide path is usually aligned with the direction of the opening end of the building, such as an L-shape or a U-shape opening. One of the building pictures from the opening end viewpoint is used as the centering position of the glide path. The second step of the landing process is to adjust the UAV's position so as to align with the glide path. This is done by matching the UAV's front camera image with the centering glide path image of the building. The partially matched result will indicate which direction the UAV should move laterally and/or vertically before an exact match is reached. At this point, the UAV has positioned itself in the glide path toward the building and is ready for the next step of landing operations.

C. Activation of Approaching

An aerial vehicle's approaching for landing is to initiate a straight ahead motion to reduce its horizontal (and vertical) distance to the target location on the ground. This phase of landing represents a target lock-on with the centering glide path image of the building. The UAV keeps moving forward while maintaining the visual centroid's fixation on the centering building image. As the approaching proceeds, the UAV will also need some image references to establish the visual lock-on. This can be done by either preparing a set of centering building image at different zoom-in levels or by extracting image features within a central zone of the centering building image. Zoom-in images are used by the object/target image feature matching approach, while the central zone image features is used for the object/target tracking approach. Both can provide horizontal and vertical guidance to reach a proper hovering point by visual navigation. This proper hovering point is on top of the designated landing spot and is away from the building's exterior wall at a minimum safe distance.

D. Descent and Touchdown

The final step of the landing process is to start a descent from the hovering point until a touchdown on the designated landing spot is achieved. Since the hovering point is vertically on top of the landing spot, the descent operation can be straightforward by giving a constant negative velocity in the z-axis to gradually decrease the altitude. In addition, a visual navigation approach with a vertical shot camera can also enhance the precision and safety of the actual landing. Similar to the visual navigation performed in the approaching step, both image feature matching and object/target tracking can be adopted to engage in a visual lock-on with the landing spot marker. This will provide guidance of minor lateral movement so as to achieve a precision landing. Unexpected obstruction of the landing spot marker can also be detected. If this occurs, the descent operation will be stopped and a landing recovery procedure initiated.

III. Autonomous Vision-Navigated Landing System Architecure

To achieve a high level of autonomy for precision landing on partially surrounded accessory floor, we adopt several technical components to construct an autonomous system capable of performing a complete navigation process for precision landing. The functional division of the navigation process includes target/object image matching, mapping and localization, navigation control, and error recovery. We adopt two types of state-of-the-art vision-based methods, image feature matching and simultaneous localization and mapping, while developing other control aspect functions. All functional components are integrated with behavior tree as the control architecture for real time execution.

A. Object/Target Image Matching

The first step of initializing a precision landing process is to check for the correct building site with stored images of designated location. For the functional need of object/target identification, we adopt the image feature matching techniques. We use the SIFT (Scale Invariant Feature Transform) feature detector from OpenCV library for our functional component. However, in some cases of regional textural similarity, SIFT is prone to erroneous matching. To reduce mismatches, we further use the RANdom Sample Consensus (RANSAC) algorithm [12], a method that exploits spatial relationships of the matching points to eliminate mismatches, for enhancement [13]. Figure 1 shows an image matching result of SIFT at the target building site with some matching errors in red circle and a better matching result with the enhancement by RANSAC. This object/target image matching function is to fulfill the landmark recognition step of the precision landing process.

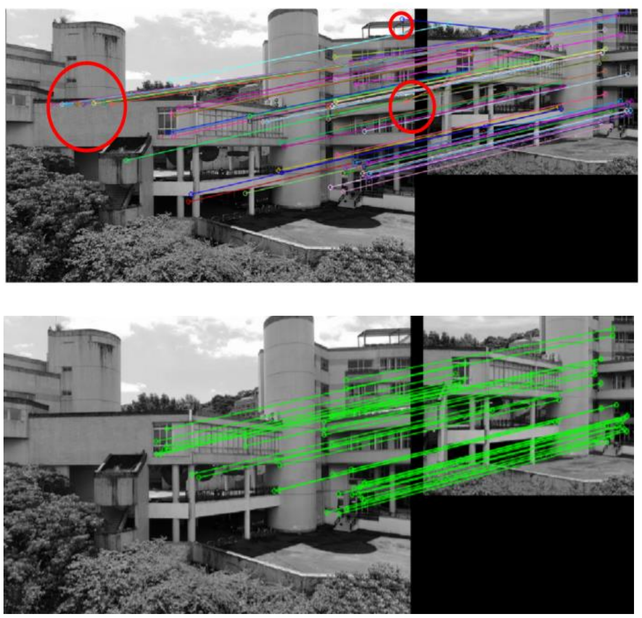


Figure 1. Object/target image matching with SIFT and by RANSAC enhancement
(Top: before enhancement, erroneous matching in red circle)
(Bottom: after enhancement, better matching result)

Once the target building site is confirmed by successful image matching result, the next step of precision landing is to orient the drone to the centering position of a predefined glide path. The object/target image matching function is also used to provide a proper guidance. To establish a proper orientation for glide path, a central portion of the drone's current image frame is used to match with the stored centering glide path image of the building. If the image matching is successful, the current orientation of the drone is accepted for approaching activation. Otherwise, position adjustment is followed by identifying regions of matched key points between the drone's current image frame and the centering image. We divide the drone's image frame into four partitions and match each image partition with the centering glide path image. The most successful matching result among the four partitions indicates a proper movement direction to be aligned with the pre-defined glide path. Figure 2 shows a scenario in which the two steps of establishing the orientation for the glide path are executed. In Figure 2 top, the first step of matching the central region of the drone's image frame with the centering glide path image is not successful. Therefore, in Figure 2 bottom, the four partitions of the drone's image frame are separately matched with centering glide path image and the upper left partition is matched most successfully. This indicates that the drone should move toward the upper left direction. A proper distance of movement to result in a view shifting between partitions can also be calculated. With proper moving direction and moving distance, the next round of image matching will successfully match the centering glide path image with the central region of the drone's image frame. Drone's motion adjustment is performed by the navigation control function. This completes the step of orienting the drone to establish a proper glide path.

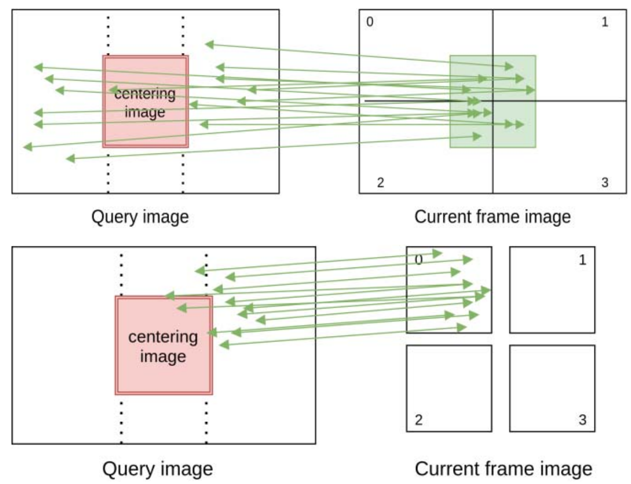


Figure 2. Image matching between a centering glide path image and the drone's current image frame for orientation adjustment

B. *Visual Simultaneous Localization And Mapping*

In order to successfully and safely navigation itself around in a partially obstructed space, the drone would need to rely on a map of the environment and a way to position itself in the map. The map would lay out perceivable objects as scene feature points. By knowing its continuously updated location on the map, the drone can dynamically calculate a suitable path toward the hovering point near the target building so as to perform the final descent on the landing spot. Considering the limited resources of computational power and sensor payload on a drone platform, we use ORB-SLAM [14] for its functional versatility and sensory simplicity. ORB-SLAM requires only monocular camera and achieves accurate results at reasonable computational cost. Scene feature points are derived by ORB features which are computationally efficient and mostly invariant to viewpoint and illumination.

One of the inherent problems is the scale ambiguity caused by the lack of depth information [15]. We use drone's IMU data to transform ORB-SLAM map scale to real-world coordinate scale. A scale ratio is derived by comparing an IMU measured moving distance in the real world with the ORB-SLAM measured pixel distance. The drone can initiate a vertical motion to obtain an estimated scale ratio between ORB-SLAM coordinates and real world coordinates. ORB-SLAM is used during the drone's approaching toward the target building until a hovering point on top of the landing site is reached. With monocular image of the target building, ORB-SLAM provides an estimated pose (position and orientation) of the drone with respect to the detected features in the image scene. This pose estimation is then used by navigation control function to conduct the pre-specified approaching process. Similarly, ORB-SLAM is also used in the final step of descent and touchdown.

C. Navigation Control

Navigation control function is responsible for conducting a continuous control cycle of giving out a motion command, measuring the motion results, comparing the motion results with the desired results, and determining an appropriate adjustment to reduce the difference. In an ideal or simulated world, a pre-planned flight path can be followed perfectly by executing the corresponding translational and rotational movement. However, wind conditions and less precise motor driven movement in the real world cause constant flight deviation. Navigation control performs dynamic correction on drone position in order to rectify occurring deviation from the pre-calculated path or the desired location. At each tick of navigation control cycle, motion results are provided by perception functions such as image matching and SLAM and a new adjusting motion is executed. In short, this realizes a real-time control system of perception, decision, and action in closed loop.

D. Behavior Tree

An autonomous mobile system operates in real world applications must be built in a mechanism to conduct a continuous cycle of perception-decision-action so as to engage with the environments for task activities. The perception-decision-action cycle repeats a process of perceiving the environment, making decision by selecting an option, and executing an action. We use the behavior trees control architecture as the perception-decision-action mechanism. Behavior trees has been a popular control architecture for artificial agents first in computer game community, and later, in robotics [16]. With a tree-like representation, a behavior tree embeds a logical structure to switch among behaviors under various conditions. Modularity and reactiveness are two attractive properties of behavior trees for developing an autonomous agent with increasing complexity in a situated environment. For our purpose of designing and controlling an autonomous drone for precision landing, behavior trees provide a good structural representation of mixing two layers of perception-decision-action and realize an effective switching mechanism in real-time execution. We have implemented a behavior tree designed for precision landing on accessory building floor and run it on a drone platform with actual flight test. The precision landing behavior tree successfully integrates and embeds functions of the technical components to provide a coherent execution toward task completion. An autonomous precision landing behavior is fully demonstrated.

IV. Experimental Results

We conduct a set of experiments to verify system autonomy and evaluate task performance. The actual system development includes a number of software and hardware components. We use DJI MAVIC 2 ZOOM, a quadrotor with internal IMU, basic latitude sensor, fish-eye front camera, Wi-Fi module, and ROS-compatible SDK, as the drone platform. A Wi-Fi link connects system software with the DJI MAVIC 2 ZOOM drone in real time. Camera image and status data on drone are sent back to system software, while flight commands are sent to drone. Image resolution is set to 640 x 368 pixels in order to provide a video stream of 30 frames per second.

The precision landing task target is a hillside building on a university campus. The building, as shown in Figure 3, has an L-shaped opening with a small patch, roughly 17 meter by 18 meter, of accessory floor. As a designated landing spot, we put a 40 cm x 40 cm AprilTag marker on the floor about 4 meter away from the building's exterior wall. The opening direction of the accessory floor is also surrounded by tree tops, which would present an obstruction to a low altitude approaching. This testing site represents a high-risk landing location for GPS-based navigation. The small-sized flat ground allows only limited margin of errors. Inadequate landing precision, caused by either perception capability or navigation autonomy, will potentially pose a safety hazard to the drone and its package. For conducting the experiments, we focus on the final landing process after GPS-waypoint navigation near the landing site. At the beginning of each experiment, the DJI MAVIC 2 ZOOM is manually control to reach at an aerial position about 40~45 meters away from the building and 6~8 meters above the building accessory floor. Once the DJI MAVIC 2 ZOOM is set to hovering, the experiment begins by switching to autonomous system control.

Figure 3. The building and its accessory floor for precision landing site

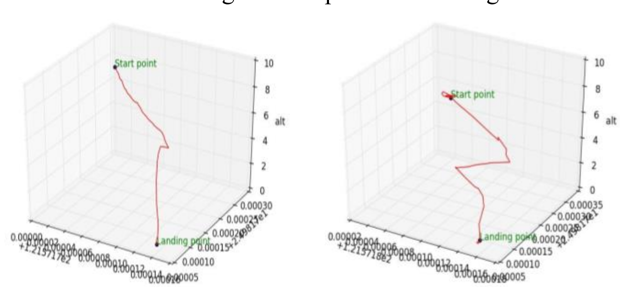
The experiments include a set of 10 autonomous flights for final landing on the designated landing spot. The 10 runs of actual flights were spread over day times in three separate days for various lighting conditions. For each flight, the initial location is set with a random deviation of 3 ~ 5 meters in the x-y-z axes. This is to create diverse scenarios of final landing in which the drone will begin with different views of the target building at various angles. The purpose is to provide a realistic test for the robustness of the autonomous system and to establish an indicative level of performance. The overall performance index is to observe whether a precision landing task can be successfully completed from landmark recognition, orientation for glide path, activation of approaching, to descent and touchdown. A distance metric is also used to measure how close the actual landing location is to the center of the landing marker.

Experiment results reported in Table 1 shows that the visual navigation approach achieves successful landing performance in all 10 trials. The average deviation distance between the final landing location and the center of the landing spot marker is 15.2 cm. This indicates that most landings are located on the 40 cm x 40 cm marker area, except for two landings just missing the marker area with deviation distances of 22.5 cm and 27.5 cm, respectively. The flight times for final landing range between 3m04s and 4m22s, reflecting the dynamic nature of real-time navigation control in position and orientation adjustment in different lighting and wind conditions. However, there is no relation between flight time spent and the resulted deviation distance. This precision landing performance is not satisfactory but may be sufficient for practical use. One of the primary factors in affecting the landing precision is the current basic level of visual-servo technique, which lacks the required precision in position control. Another source of position errors may come from the camera location of the DJI MAVIC 2 ZOOM, which is at the front area of the drone body.

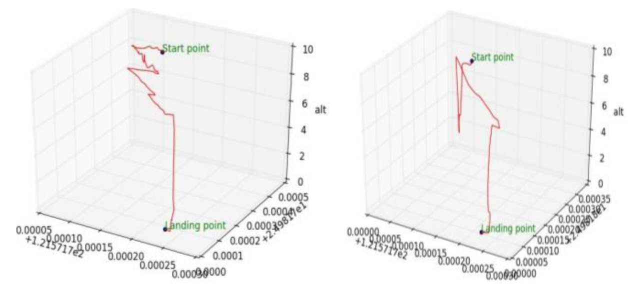
TABLE 1. PERFORMANCE ON PRECISION LANDING

Round of Flight Test	***Distance to marker center***	***Flight Time***
1	22.5 cm	3m51s
2	14.3 cm	4m22s
3	8.6 cm	3m23s
4	15.4 cm	3m44s
5	12.5 cm	3m50s
6	27.5 cm	3m32s
7	19.4 cm	3m18s
8	10.0 cm	3m25s
9	10.0 cm	3m04s
10	11.8 cm	3m42s
Average	15.2 cm	3m37s

For further detailed observation, we track the trajectories of actual landing flights by using GPS coordinates record. Figure 4 shows four trajectory instances from a starting point in space for autonomous landing navigation to the final landing spot. This provides a visual reference to the dynamic process of navigation control in which the drone's positions are adjusted based on visual recognition of the target images. Overall, the trajectory observation shows that, despite not being perfect, an autonomous visual navigation for precision landing is achieved.



(a) (b)



(c) (d)

Figure 4. Trajectories of autonomous visual navigation for precision landing

V. CONCLUSION AND FUTURE WORK

This work presents a technical approach to develop a vision-based autonomous navigation system for precision landing on building accessory floor. We use behavior tree to integrate various functional modules, from object image matching, ORB-SLAM, to navigation control. The developed system has been implemented on a quadrotor and field tested on a real building accessory floor in an outdoor environment. Initial experiments, including a total of 10 rounds of actual flights, show successful system autonomy and acceptable task performance. This overall achievement leads us to believe that our technical approach has been validated and can be universally applied to common multi-rotors. Based on this initial success, we plan to conduct further technical improvement and experimental validation. On the technical side, we will work on more sophisticated visual-servo technique and position calibration with the non-central downward viewing camera. On the experimental side, more field test scenarios, including various landing sites at different building locations, are also considered.

ACKNOWLEDGEMENT

This research is partially supported by the Ministry of Science and Technology under grant number 109-2634-F004-001 through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan.

REFERENCES

[1] Paul Groves, Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition, Publisher: Artech House, 2013.

[2] K. Peng, A secure network for mobile wireless service, Journal of Information Processing Systems, vol. 9, no. 2, pp. 247–258, 2013.

[3] P. A. Zandbergen and L. L. Arnold, Positional accuracy of the wide area augmentation system in consumer-grade GPS units, Computers and Geosciences, vol. 37, no. 7, pp. 883–892, 2011.

[4] Y. Feng, C. Zhang, S. Baek, S. Rawashdeh, and A. Mohammadi, Autonomous landing of a UAV on moving platform using model predictive control. Drones, 2(4), 34, 2018.

[5] A. Cesetti, E. Frontoni, A. Mancini, P. Zingaretti, and S. Longhi, A vision-based guidance system for uav navigation and safe landing using natural landmarks, in Selected papers from the 2nd International Symposium on UAVs, Reno, Nevada, USA June 8–10, 2009. Springer, 2010, pp. 233–257.

[6] D. Lee, T. Ryan, and H. J. Kim, Autonomous landing of a vtol UAV on a moving platform using image-based visual servoing, in Proc. IEEE Int. Conf. Robot. Autom., 2012, pp. 971–976.

[7] H. Choi, M. Geeves, B. Alsalam, and F. Gonzalez, Open source computer-vision based guidance system for UAVs on-board decision making," IEEE Aerospace conference, Big sky, Montana, 2016.

[8] C. Forster, M. Faessler, F. Fontana, M. Werlberger, and D. Scaramuzza, Continuous on-board monocular-vision-based elevation mapping applied to autonomous landing of micro aerial vehicles. In Proceedings of the IEEE International Conference on Robotics and Automation, Seattle, WA, USA, 26–30 May 2015, pp. 111–118.

[9] P. H. Nguyen, M. Arsalan, J. H. Koo, R. A. Naqvi, N. Q. Truong, and K. R. Park, LightDenseYOLO: A fast and accurate marker tracker for

autonomous UAV landing by visible light camera sensor on drone, Sensors, 18, 1703, 2018.

[10] J. Wubben, F. Fabra, C. T. Calafate, T. Krzeszowski, J. M. Marquez-Barja, J. C. Cano, and P. Manzoni, Accurate landing of unmanned aerial vehicles using ground pattern recognition, Electronics, 8, 1532, 2019.

[11] S. G. Lin, M. A. Garratt, A. J. Lambert, Monocular vision-based real-time target recognition and tracking for autonomously landing an UAV in a cluttered shipboard environment, Autonomous Robots, 41(4): 881–901, 2017.

[12] M. Fischler and R. Bolles, Random sample consensus: A paradigm for model fitting applications to image analysis and automated cartography, Proc. Image Understanding Workshop, pp. 71-88, 1980-Apr.

[13] H. Zhou, Y. Yuan, and C. Shi, Object tracking using SIFT features and mean shift, J. Comput. Vision Image Understand., vol. 113, no. 3, pp. 345–352, Mar. 2009.

[14] R. Mur-Artal and J. D. Tardos, ORB-SLAM2: an Open-Source SLAM system for monocular, stereo and RGB-D cameras, arXiv:1610.06475, Oct. 2016

[15] S. Choi, J. Park, W. Yu, Resolving scale ambiguity for monocular visual odometry, in Proc. IEEE URAI, pp. 604-608, 2013.

[16] M. Colledanchise and P. Ögren, Behavior trees in robotics and AI: An introduction. CRC Press, 2018.