Implementation of a Monocular ORB SLAM for an Indoor Agricultural Drone

2023 Third International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP) | 978-1-6654-7353-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICA-SYMP56348.2023.10044953

Kanjanapan Sukvichai
Dept. of Electrical Engineering
Kasetsart University
Bangkok, Thailand
fengkpsc@ku.ac.th

Noppanut Thongton
Dept. of Electrical Engineering
Kasetsart University
Bangkok, Thailand
noppanut.th@ku.th

Kan Yajai
Dept. of Electrical Engineering
Kasetsart University
Bangkok, Thailand
kan.y@ku.th

***Abstract*— Drones are increasingly being used in almost every major industry, including agriculture. Intelligent drone systems could enable precise agricultural. One of the critical uses of agricultural drone is to use in an automatic plant monitoring and inspecting. A drone must be tiny enough to fly between plants in order to capture images of plant trees or fruits in an indoor environment. Therefore, drone's payload is crucial because it limited onboard sensors weight. SLAM is necessary for autonomous navigation because it could provide all necessary information of drone navigation system without collisions. ORB SLAM was popular for monocular systems since it extracted ORB features from images to generate Visual Odometry. ORB SLAM generated the map as the output that can be used to estimate the position of the drone without the assistance of other sensors. In this research, monocular ORB SLAM system was proposed, explained and experimented in order to obtain and confirm the ORB SLAM performance for an indoor application. The experimental result showed that the ORB SLAM worked properly for generating a map of a tomato greenhouse and the output map could be used to estimate the drone position with some limitation.**

***Keywords*—Agricultural drone, ORB SLAM, ORB features, position estimation, indoor drone.**

I. INTRODUCTION

Small Autonomous Unmanned Vehicles (AUVs), particularly drones, are increasingly being used in essentially major sector of the economy, including agriculture. The agricultural drone market is expected to grow from $1.2 billion (USD) in 2019 to $4.8 billion (USD) in 2024, according to a forecast [1]. In the coming years, drones will become more widespread on both large and small-scale farms. Several drone technologies are already being deployed on farms, while some new agricultural drone technologies are being investigated. Agricultural drone applications included scouting or monitoring plant health using visible and near-infrared color spectrums, inspection of large-scale crop field conditions for estimating yield, planting or seeding, pesticide and fertilizer spraying, and irrigation. Drones were the key of the precision agriculture especially in a greenhouse. Due to a lack of human employees, autonomous drones are fascinating and significant in agriculture in the near future. One major challenge for an autonomous drone works in a greenhouse is maneuverability because the area was clustered by large amount of plants. Drone must be small enough to maneuver between plants but has enough calculation power to collect and preprocess necessary information for a navigation.

A tomato is a valuable plant for human because its nutrition. It was commonly found and used as the main component in food preparations across the world. A tomato is simple to grow, but even though it was grown in a contained environment, it was easily plagued by infections. Disease is the most common cause of decreased product and plant quality, which affects their worth. Faster detection will assist farmers in controlling the harm that may occur. Disease detection is a difficult undertaking for farmers who must observe disease across a vast area. As a result, the use of drones to detect a disease in a field is autonomous, reducing a difficult duty for farmers while increasing product quality, resulting to better prices.

In this research, the Simultaneous Localization and Mapping (SLAM) is focused and required for autonomous applications. SLAM algorithm requires a specific sensor, such as a 3D-Camera or Light detection and ranging (LIDAR), to generate a map for navigation. The most significant issue with drone implementation is a low payload limitation, thus, it cannot carry many sensors onboard. Oriented FAST and Rotated BRIEF SLAM or ORB SLAM [2] is an effective method that requires only a monocular camera to construct a 3D map of the plantation area and the created map can be used for autonomous navigation for inspecting the anomalies of plant. Gazebo, a virtual environment with a physics engine, is highly useful for testing drones and algorithms before they are tested in the actual field. A monocular ORB SLAM will be explored, evaluated and experimented in simulation in this research.

II. ORB SLAM

Simultaneous localization and mapping or SLAM is the computational problem of creating and updating a map of an unknown environment while simultaneously tracking a robot's location. Oriented FAST and Rotated BRIEF SLAM or ORB SLAM is a popular visual based SLAM and is a versatile and accurate SLAM algorithm that can compute the camera trajectory and a sparse 3D reconstruction of the scene in a wide variety of environments in real-time based on Visual Odometry for feature extracting and tracking. ORB SLAM for agriculture application was evaluated in the previous research [3]. In this topic, Visual Odometry and ORB SLAM were explained.

A. Visual Odometry

Odometry is the process of utilizing motion sensor information to estimate position change over time. It is widely used in robot systems to calculate their position relative to a starting point. Visual Odometry (VO) basically is the Odometry system that used cameras as the sensors to estimate the posture and location. There are 2 major types of Visual Odometry which are a monocular and stereo VO. The monocular VO uses a single camera while stereo VO uses multiple cameras. There are advantages and disadvantages from both stereo and monocular VO. The advantage of stereo is that it allows you to estimate the exact trajectory, whereas monocular just allows you to estimate the trajectory, which is unique up to a scale factor. Therefore, in monocular VO, it provides information of motion in unit such as one unit in x, two units in y, and three units in z axes, while in stereo VO, it provides information of motion in real dimension such as one meter in x, two meters in y, and three meter in z axes. In this research, due to the limitation of drone payload and onboard processor computation power, only one camera can be installed to a drone system. Thus, monocular VO is the main focus in this topic. The Visual Odometry could be achieved by features matching between two consequent video frames based on key-points detector and descriptor. Scale-Invariant Feature Transform or SIFT [4] was the well-known key-points detector and descriptor. SIFT has been successful in a variety of applications that involve usage visual features. However, it entails a significant processing load, particularly for real-time systems like visual odometry or low calculating power devices like onboard computer in a drone system.

Oriented FAST [5] and Rotated BRIEF [6] or ORB algorithm has similar matching performance to SIFT algorithm but less affected by image noise and is capable of being used for real-time performance. ORB algorithm first finds key-points using FAST, then uses the Harris corner measure to find the top N points among them, then, employs a scale pyramid of the image to generate multiscale features. In order to obtain orientation information, ORB computes the intensity weighted centroid of the patch with the corner located in the center. ORB uses steer BRIEF descriptor that related to the orientation of key-points. The key-points from two consequent video frames were shown in Figure 1.

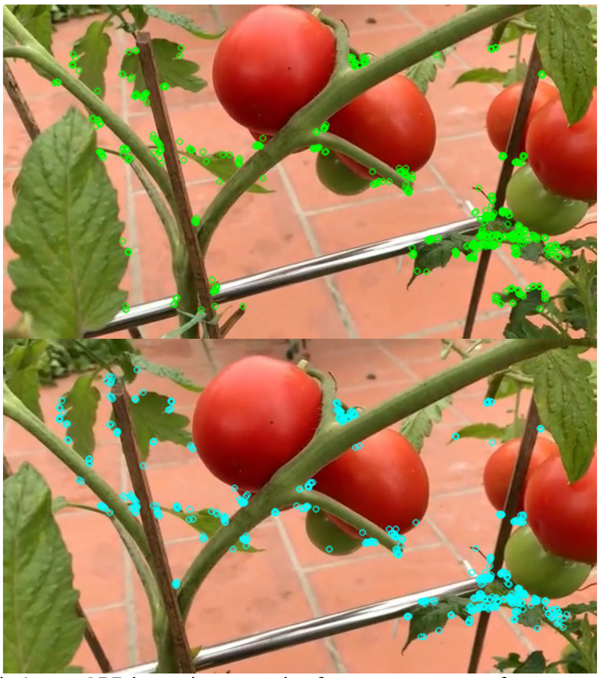

Fig 1. ORB key-points extraction from two consequent frames.

By considering the key-points from two frame, the Visual Odometry could be obtained by feature matching algorithm. There are many matching algorithms such as Brute-Force and FLANN [7] Matcher. Both algorithms used the K nearest neighbor and norm approach for matching key-points from two frames. The result from FLANN ORB key-points matching result is shown in Figure 2. From the result, the closest key-points from two frame were matched.

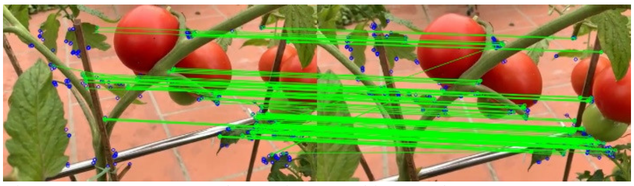

Fig 2. FLANN ORB key-points matching result.

Homograph of two frames could be obtained after matching process. Homography is a transformation that maps the points in one point to the corresponding point in another image. Therefore, the distance and orientation between two frames can be calculated and creates the Visual Odometry. Homography matrix of result shown in Figure 2 is shown in (1). Due the limitation of monocular, it is clear that the Homography cannot be used to obtain the translation in x, y and z axes in the real unit. Only the angles different between two frames could be calculated via SVD process. The roll angle of two frames shown in Figure 2 is 0.93842 degree.

$$H = \begin{bmatrix} 1.0494 & -0.0257 & -26.5705 \\ 0.0126 & 1.0263 & 14.7974 \\ 0.0000 & 0.0000 & 1 \end{bmatrix} \quad (1)$$

B. ORB SLAM

In general, SLAM that uses image features capture from mono camera, stereo camera and RGBD camera to localization and mapping is called Visual SLAM or V-SLAM. There are many V-SLAMs such as RTAB-Map, Open VSLAM and ORB SLAM. ORB SLAM is significantly more accurate, especially when the same location is revisited. ORB SLAM's precision is derived by non-linear bundle adjustment, which involves deriving observations of the same map location from widely separated key-points in frames. The important frame that contains useful information is called the keyframes. ORB SLAM can detect key-points matches in keyframes even when they are widely separated in time. Bundle Adjustment (BA) is known to provide accurate estimates of camera localizations as well as a sparse geometrical reconstruction [8]. ORB SLAM consists of many states in order to meet necessary requirements to generate map and localization simultaneously including ORB features matching, DBoW2 place recognition, pose graph optimization, local BA, global BA, and map management. The flow of ORB SLAM is shown in Figure 3 and the output points cloud map is shown in Figure 4. The points cloud map is represented in three dimensions which could be used for localization and navigation system in order to navigate drone in known environment after map is collected.

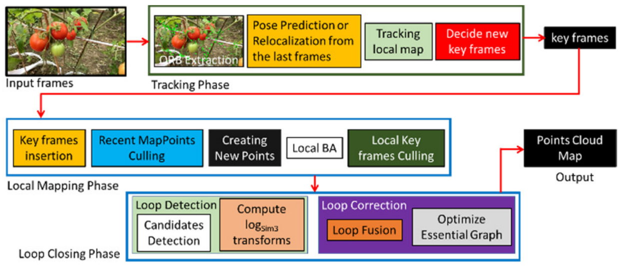

Fig 3. ORB SLAM process.

III. Drone System

Greenhouse drone is designed based on a quadcopter structure. It was a compact drone that can maneuver in a compact tomato farm. The robot span is 80 cm with weight less than 1 kg. Four rotors were installed to drone frames. Drone body constructed from composite materials including balsa wood and carbon fiber. Raspberry Pi4 was installed to the drone as the main computer. One C920 Logitech webcam was attached to the drone and connected to Raspberry Pi4 board. Pixhawk PX4 PIX 2.4.8 32 Bit flight controlled was installed and connect to Raspberry Pi4 via Robot Operating System (ROS).

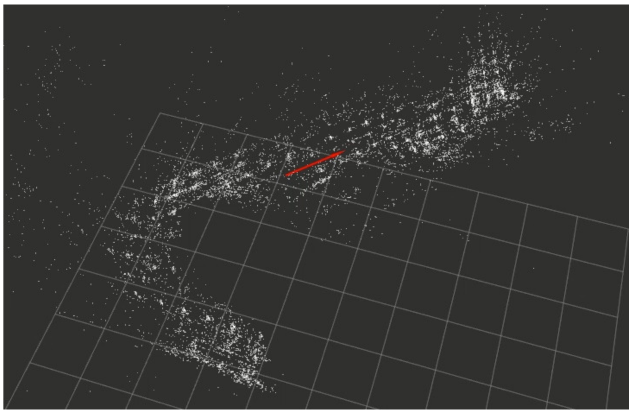

Fig 4. Points cloud map output from ORB SLAM.

The designed greenhouse drone is shown in Figure 5. Ubuntu 20.04 operating system was installed in the Raspberry Pi along with ROS Noetic. Drone can send useful information to base computer via 2.4 GHz wifi connection. The designed drone had ability to mobilize through a small area, hovering in specific high and location, strong enough to withstand falls and light weight. Figure 6 shows a test flight of a developed drone.

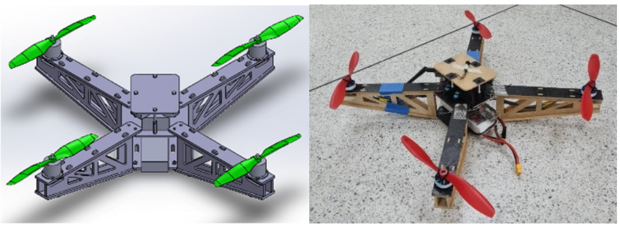

Fig 5. Designed greenhouse drone: Model (Left) and Real (Right)

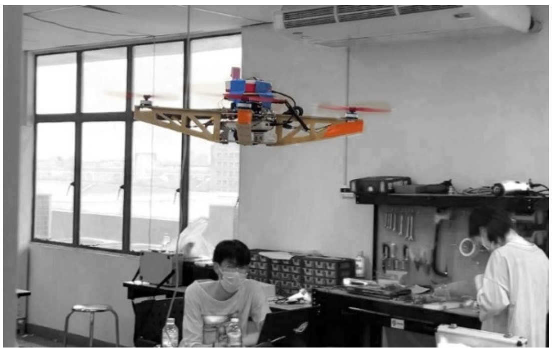

Fig 6. A test flight of a developed drone

IV. Experiments

The experiments were conducted in order to obtain the performance of the drone and ORB SLAM. First experiment was conducted for testing ORB SLAM performance. The large greenhouse video was used to evaluate ORB SLAM. The experiment was conducted in RVIZ under ROS environment. The experiment is shown in Figure 7.

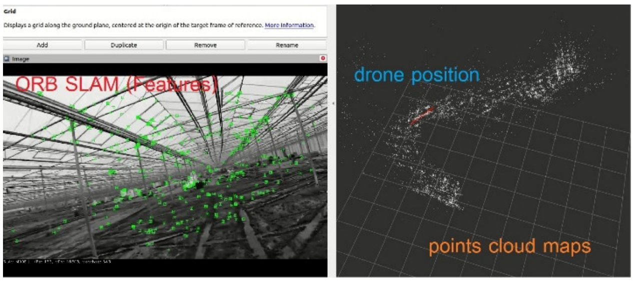

Fig 7. ORB SLAM testing result from a large greenhouse video.

From the experiment result, ORB SLAM can generate a proper points cloud map. The drone position tracked by ORB SLAM map is shown in Figure 8. This result confirm that the output ORB SLAM could be used for drone's localization and navigation.

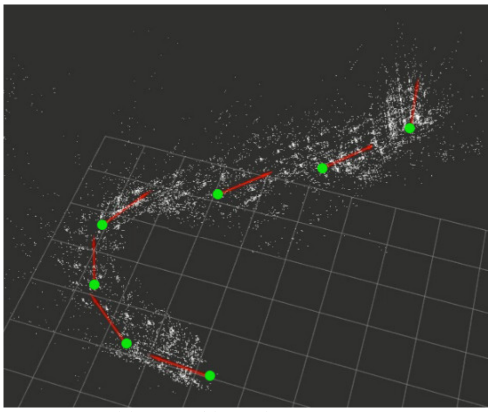

Fig 8. Drone positions tracking using ORB SLAM map.

Gazebo [9], a virtual environment with a physics engine, is highly useful for testing drones and algorithms before they are tested in the field. Gazebo is easy to be connected to ROS environment and it works well with ROS publish subscribe communication. Drone Unified Robot Description Format (URDF) file that describes the physical structure of the drone was added into Gazebo environment along with properties of propellers such as drag and lift coefficient to create the proper physical behavior in simulation. In this second experiment, ORB SLAM and drone is simulated in Gazebo as shown in Figure 9. Simulation result confirmed that the drone and ORB SLAM work properly and the real hardware experiment can be conducted in the third experiment.

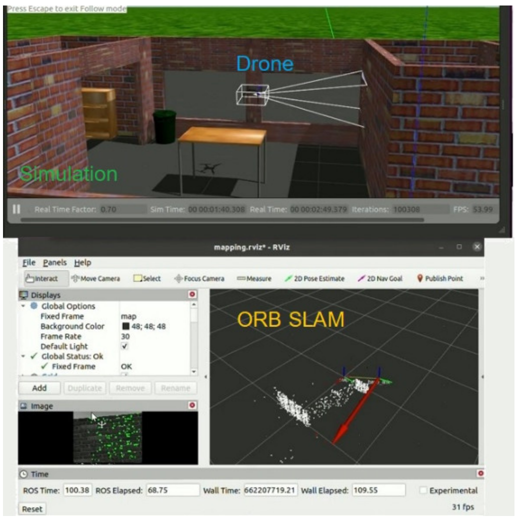

Fig 9. ORB SLAM simulation in Gazebo environment.

The third experiment was conducted by the real drone in two scenarios. The first scenario was tested by flying a drone in an indoor room and ORB SLAM map was created. The collected map was evaluated by flying the drone in the same environment again in straight line and the position of the drone obtained by ORB SLAM map showed straight line as the output drone trajectory. The first scenario experiment is shown in Figure 10.

Fig 10. Drone flying indoor in straight line in the frist scenario.

The drone is tested again in the same environment with waypoints that are the square's corners in the second scenario. Figure 11 displays the position of the drone in the second scenario, while Figure 12 displays the error of drone position from Intel RealSense T265 in the x and y axes. To compare with the ground truth, the drone location from ORB SLAM was multiplied by the ratio that determined by distance from the real world divided by distance from ORB-SLAM. The ORB SLAM works properly according to the experimental result. The position inaccuracy occurred, and it was resulted from the monocular method's constraint that only one camera cannot measure the true unit of the environment in meters. When the input frame is blurred due to drone motion, the error additionally occurred due to inadequate feature extraction from the ORB algorithm. The experimental result was acceptable for using in the real application since the error could be reduced by adding extra features such as QR sticker as reference points in the field.

V. Conclusion And Future Work

Agricultural drones are becoming increasingly significant in both small and large-scale agricultures. In a small indoor farm such as tomato farm, a drone must be small enough to fly precisely between plants limited space. A small drone has limited payload and it can carry limited sensors. Therefore, one camera can be installed into the drone system in this research. ORB SLAM was proposed in this research due to its performance and it worked well with monocular system. ORB features could be extracted from the tomato plants, fruits and vine images captured from drone's onboard camera. The features were used to estimate Visual Odometry of the drone without the assistance of other sensors. The experiments were conducted to obtain performance of the monocular ORB SLAM drone system. The experimental results showed that the ORB SLAM cloud create the proper map that can be used to estimate drone position and can be used to navigate a drone to already known environment. Position error happened due to limitation of the monocular system, but it is still good enough to use for navigating a drone with limited payload in a small area. The future work of this research is to test the proposed monocular ORB SLAM drone system in iMAV2022 real tomato farm competition.

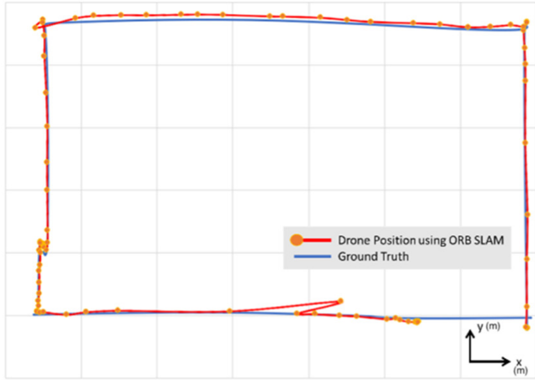

Fig 11. Trajectory of drone compared with the ground truth.

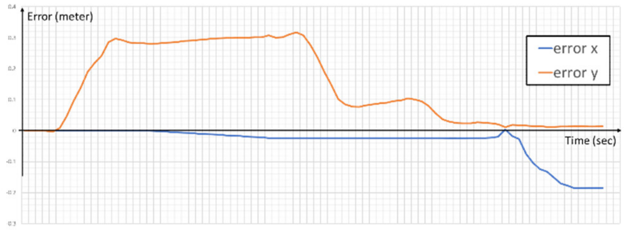

Fig 12. Position error of the drone from the second scenario.

References

[1] Agriculture Drones Market with COVID-19 Impact Analysis, by Application, Offering, Farming Environment, Farm Produce, Component And Geography - Global Forecast to 2025, [Online]. Available: https://www.reportlinker.com/p03903671/?utm_source=GNW

[2] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015, doi: 10.1109/TRO.2015.2463671.

[3] S. Krul, C Pantos, M. Frangulea and J. Valente, "Visual SLAM for Indoor Livestock and Farming Using a Small Drone with a Monocular Camera: A Feasibility Study," Drones 2021, 5, 41. doi:10.3390/drones5020041

[4] D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the International Conference on Computer Vision. Vol. 2. pp. 1150–1157, 1999, doi:10.1109/ICCV.1999.790410

[5] R. Edward and T. Drummond. "Machine Learning for High-Speed Corner Detection," Computer Vision ECCV 2006, 7 May 2006, doi: 10.1007/11744023_34

[6] M. Calonder, V. Lepetit, C. Strecha and P. Fua, P. "BRIEF: Binary Robust Independent Elementary Features," In: Daniilidis, K., Maragos, P., Paragios, N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol 6314. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-15561-1_56

[7] M. Muja and D. G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration," in International Conference on Computer Vision Theory and Applications (VISAPP'09), pp. 331–340, 2009.

[8] B. Triggs; P. McLauchlan; R. Hartley; A. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," ICCV '99: Proceedings of the International Workshop on Vision Algorithms. Springer-Verlag. pp. 298–372. doi:10.1007/3-540-44480-7_21

[9] [Online] https://gazebosim.org/home