Loacalization of Vision assisted Navigation for UAV: A brief Survey

Iftekharul Islam Shovon and Seokjoo Shin*
Dept. of Computer Engineering
Chosun University,
Gwangju, South Korea
*Corresponding author: sjshin@chosun.ac.kr

***Abstract*—Vision-assisted UAV navigation is not a new topic. Many applications require the help of vision-based systems, and navigation is one of the main focuses, along with vision-based car automation. Vision-based navigation opens up many opportunities in both the research and application field. In this paper, we briefly surveyed the localization techniques of vision-based UAV, categorized them into two categories, mapped and mapless and further subclassified them into three classes, explored the critical ideas of existing systems, discussed their advantage and disadvantages, and at the end, pointed out some research ideas for future work.**

***Index Terms*—UAV, Vision assisted navigation, localization,**

I. Introduction

Navigation is the process of reaching a destination from a specific starting point safely and quickly. UAV navigation serves the same purpose. Hence, UAVs must avoid obstacles and create a path according to their environment to reach the destination. Traditionally there are three ways of navigating UAVs, GPS-based navigation, Inertial Navigation System (INS), and Digital Terrain System (DTS). GPS, in theory, gives the most precise location information, but in practice, it lacks usability in dense forests, mountains, or if the signal is jammed. The alternative to GPS is DTS, which records the height from the ground, matches the information with a pre-stored digital map, and locates the position. However, if the size of nearby places is the same, the system finds it difficult to identify the exact location. On the other hand, INS is widely used for commercial aviation flights and is one of the first UAV navigation techniques. It requires initial coordinates (IN) to initiate. It incorporates velocity and acceleration to find the location. Despite being popular, it suffers from error accumulation, and the error increases with time. The alternative to this traditional navigation system is vision-based navigation. The data is taken directly in the form of images. It estimates the current position and obstacles using frames, which eliminates IN problem. Bad weather may cause damage to the camera, but the advantages suppress the drawbacks making it much more attractive to the research community. In this paper, we briefly discuss the process of Vision-based navigation and compare the existing models and compare each other. Section II describes the process of Vision-based navigation and compares them, and we end the paper with open challenges and a conclusion in sections IV and V, respectively.

II. Vision Based Navigation

Vision sensors provide more detailed information about the surroundings than traditional sensors like GPS and Ultrasonic sensors and are cheaper than conventional sensors. Figure 1. shows the flowchart of the process of Vision-based navigation. Typical sensors for vision-based include a) Monocular, b) Stereo, c) RGB, and d) Fisheye cameras. A monocular camera is a common type of vision sensor. The most common use for monocular cameras is finding the minimum weight, and compactness is essential.

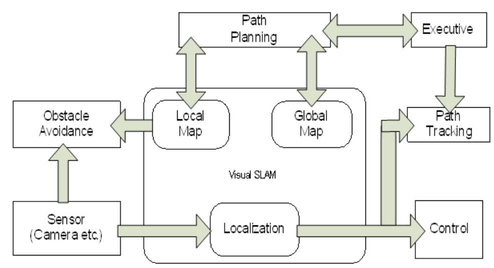

Fig. 1. Flowchart of vision-based navigation

One of the most crucial tasks of vision-based navigation is localization and mapping. Authors of [1] classified localization and mapping into three types: a) Map-based, b) Map-building, and c) Map-less. This paper focuses on Map-building localization techniques.

A. Map-Building

It becomes difficult to navigate using a preexisting map during or after any disaster. Hence, building a map of the environment while flying is an attractive solution to overcome this limitation. As UAVs get smaller, traditional sensors like laser radar are fazed out, favoring tinier vision-based sensors. The rapid development of visual simultaneous localization and mapping (SLAM) techniques has aided in using the map-building method. Researchers derived three different methods for map-building, a) Indirect Method, b) Direct method, and c) Hybrid method.

- Indirect Method: Indirect map-building detects the features of the image as input for motion estimation and localization. For more than a decade, scientists have studied various types of feature detection, and most of the current SLAM algorithms fall under the feature detection framework. In [6], Davision proposed revolutionary work on a top-down Bayesian framework for a single camera module for localization. The result shows the real-time mapping of the environment using a monocular camera and SLAM algorithm. On the other hand, [2] proposed a parallel mapping and tracking algorithm process; this is used as a standard for most SLAM systems. [3] proposed a loop detection closure algorithm for large-scale navigation. Authors in [4] proposed a scheme for indoors where the trajectory is unknown to navigate without the help of GPS using a single camera. They used energy-based feature points and straight-line architecture for navigation. On the other hand, authors of [5] proposed a multicamera parallel tracking mapping using ego-motion estimation from different cameras to parallel process estimation and mapping.

- Direct Method: While the indirect method mentioned in the previous section performs well in most circumstances, there are limitations, especially when the environment is textureless. So, to overcome this, [6] proposed a novel idea of using images entirely and using the intensity of the picture to produce a dense map. The process is robust. Authors in [7] use the 6DOF motion of the camera to estimate the movement of the UAV using a frame rating of the entire image alignment and generate a dense map based on the feature estimation of the features. [8] uses GPU and efficiently employs a direct probabilistic approach to produce a semi-dense map. [9] takes scale factor and uses graph optimization over Sim3, allowing scale drift correction and loop closure detection. [10] proposed a navigation system based on mapped landmarks where the position of the landmarks is known. It continuously corrects the noise of the IMU image as a variable state increases the calculation time.
- Hybrid Method: The hybrid method consolidates the direct and indirect methods. It instates highlight correspondences utilizing indirect methods and consistently refines camera presents by direct methods, making it quicker and more precise. In [11], SVO proposed a semi-direct algorithm for estimating the state of UAVs, they evaluate the motion and implement cloud mapping in two different threads. They also use pixel brightness and gradient information and combine them with the alignment of the feature point for more accurate motion estimation. A high frame rate camera is needed to use SVO. Hence, it was designed for use in the onboard application and with limited computational power.

III. Open Challenges

The work on vision-based navigation is a work in progress. It has leaped forward a great deal, yet there lay problems processing real-time data from multiple sensors, and it suffers significantly from computation complexity. Vision-based UAVs require either building or having a pre-existing 3D environment map. Hence, the computational complexity is higher, and with the trend of UAVs shrinking, it becomes more challenging to process everything in a compact structure. One solution can be using FPGA-based boards to reduce UAVs' computational power and size. The problem of solving 3D mapping are NP-hard problems, and they don't have any solution; the modern algorithms also suffer from the local minimum problem, which affects the system's robustness. More research is needed in this sector. The use of multi-sensors is still inefficient, so there are many opportunities for improvement in this field. Power constraints also cripple the performances of UAVs. Due to power inefficiency, a single UAV can not perform many tasks alone. The improvements in vision-based navigation will allow multiple UAVs to work in sync, helping solve many problems.

IV. Conclusion

In this paper, we reviewed the vision-based localization process of UAVs. We classified the process into two portions, map-less and mapped. The mapped system was divided into direct, indirect, and hybrid categories. We discussed these systems in detail and compared them in Table II. Next, we discussed the future challenges of Vision-based navigation and gave further directions on achieving them.

Acknowledgment

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF2018R1D1A1B07048338).

References

[1] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[2] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces. ismar 2007: 6th ieee and acm int symp on mixed and augmented reality; 2007," 2007.

[3] I. Mahon, S. B. Williams, O. Pizarro, and M. Johnson-Roberson, "Efficient view-based slam using visual loop closures," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1002–1014, 2008.

[4] K. Celik, S.-J. Chung, M. Clausman, and A. K. Somani, "Monocular vision slam for indoor aerial vehicles," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1566–1573, IEEE, 2009.

[5] A. Harmat, M. Trentini, and I. Sharf, "Multi-camera tracking and mapping for unmanned aerial vehicles in unstructured environments," *Journal of Intelligent & Robotic Systems*, vol. 78, no. 2, pp. 291–317, 2015.

[6] G. Silveira, E. Malis, and P. Rives, "An efficient direct approach to visual slam," *IEEE transactions on robotics*, vol. 24, no. 5, pp. 969–979, 2008.

[7] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *2011 international conference on computer vision*, pp. 2320–2327, IEEE, 2011.

[8] J. Engel, T. Schöps, and D. Cremers, "Computer vision—eccv 2014," in *Proceedings of the 13th European Conference, Zurich, Switzerland*, pp. 6–12, 2014.

[9] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g 2 o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation*, 2011.

[10] H. Lan and S. Jianmei, "Research of autonomous vision-based absolute navigation for unmanned aerial vehicle," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1–6, IEEE, 2016.

[11] C. Forster, M. Pizzoli, and D. Scaramuzza, "Robotics and automation (icra)," in *2014 IEEE International Conference on*, pp. 15–22, 2014.