Visual Heading Estimation for UAVs in Indoor Environments

A.Salib
Department of Geomatics
University of Calgary
Calgary, Canada
abanob.salib@ucalgary

A.Moussa
Elec. and Comp.Dept.
Port-Said University, Egypt
University of Calgary
Calgary, Canada
amelsaye@ucalgary.ca

M.Moussa
Department of Geomatics
University of Calgary
Calgary, Canada
mohamed.moussa1@ucalgary.ca

N.El-Sheimy
Department of Geomatics
University of Calgary
Calgary, Canada
elsheimy@ucalgary.ca

***Abstract*—Recently, The UAV (Unmanned Aerial Vehicle) industry is getting a lot of attention, especially for very small drones that could fly indoors. Such a small drone can help perform rescue tasks such as investigating gas leaks or emergency situation that imposes risk on human intervention. Additionally, the UAVs can be utilized for indoor inspection and 3D mapping. The indoor environment is very challenging for the UAVs since the Global Navigation Satellite System (GNSS) cannot be reliably accessed to help UAV to locate itself. Various sensors could be mounted on UAVs to help the surveillance and navigation aspects of the operation. Range sensors such as 2D Lasers and LIDAR are very useful for providing valuable measurements towards reliable localization algorithms. Including such sensors will typically raise the system cost and limit the flight time due to increased power consumption. This research aims to assess the potential of using the typically installed UAV main camera to help estimate the UAV heading. The typical indoor environment includes many challenging situations such as monochrome surfaces and identical repeated patterns, especially in the ground surface. The research investigates the performance of different heading estimation approaches using a minimum cost configuration (without laser scanners). The proposed integration between the drone forward camera and the downward camera enhanced the navigation result compared to the other individual solutions using the downward camera only, forward camera only, or magnetometer. The performance of the investigated approaches in a real indoor flight is presented and discussed.**

Keywords*— *UAV, visual odometry, monocular camera, heading, magnetometer

I. Introduction

. While UAVs can smoothly fly outdoors with the information provided from GNSS to localize itself, they face a huge challenge to localize itself in indoor environments without GNSS information. Many applications can employ drones in indoor environments to help perform various tasks such as surveillance missions, remote inspection, search and rescue, and disaster assessment. The typical drone setup for flying in an indoor environment usually comprises a LiDAR (Light Detection And Ranging) sensor among the other sensors to help build a map for the surrounding environment and to localize itself based on the range measurements through transmitted and received laser signals.

Several enterprises are focused on developing improved drone hardware and applications, such as DJI, Hubsan, Parrot, and Yuneec. The drone market is recently expanding to include drones' weight below 250 grams [1]. The UAV regulations in different regions of the world consider drones below 250 grams as harmless drones and do not require a license to fly or registration[1]. According to rules in many regions, drones are categorized into two different types: basic drones and more advanced drones, where each type follows a different set of rules [1].

Micro Aerial Vehicles (MAVs) should typically be very close to sightline with their operator for rescue and search missions[2]. Controlling the drone in an indoor environment is extremely challenging due to the minimal space and may cause damage to the drone itself due to the expected operator stress[3]. For the reasons mentioned above, it is very desirable to fly drones in an indoor environment without any human intervention[3]. Low-cost drones can be a handy tool for many search and rescue missions. With only minimum configuration, such as an inertial measurement unit (IMU) and regular cameras, the drone can fly autonomously by connecting it to a computer [3].

This paper investigates the performance of different heading estimation approaches using low-cost configuration in indoor environments. Three approaches for heading estimation using cameras and magnetometers have been evaluated for comparison purposes.

Section II presents some related work to the current research. Section III depicts the methodology for the investigated heading estimation approaches. Section IV covers the performed test and discusses the comparison results.

II. Related work

Autonomous MAVs navigation in indoor environments depends on various sensors to compensate the inaccessibility of GNSS systems. The laser range finders such as the Velodyne sensor has a relatively high weight compared to the typical weight of such MAVs, so 2D scanners are the most commonly employed lasers with MAVs [3]. RGB-D camera is another option that could be employed for indoor navigation with its ability to capture depth measurements based on the structured light method. The depth information sensed by both laser scanners and RGB-D cameras enables more robust navigation of drones in indoor environments. The vision-based navigation has a significant potential to navigate the UAVs, especially with the help of inertial sensors; however, the navigation states could drift over time because of the accumulative fashion of state estimation between successive frames [3]. A robust technique proposed by Stefan Hrabar[4] combines stereo navigation and optical flow to guide the UAVs in an Urban area. Lippiello[5] offers another vision-based approach for obstacle avoidance in an indoor environment using optical flow and inertial measurements of MAVs. The drone's trajectory is adjusted based on the obstacle observed by the depth map obtained using the optical flow algorithm [5]. Drone visual data has been employed in [6] to support autonomous takeoff, landing, and cruise maneuvers. Using the optical flow technique, a wall collision avoidance method is proposed using the depth map information built using visual data [7]. Visual Inertial Odometry (VIO) algorithms are generally used in MAVs and offer the drone's location in a GNSS denied environment. To ensure indoor harmless autonomous navigation for MAVs via corridors, an omnidirectional fisheye camera has been utilized as the main sensor and integrated with IMU data [7].

The optical flow algorithm of downward vision has been employed successfully to assist MAVs maneuvers and navigation in indoor environments. The relative motion V between the imaging sensor and a point P is described as follows[9]:

$$V = -T_{tr} - \omega * P \tag{1}$$

where T_{tr} is the translational component of the motion
ω is the angular velocity

And the relation between the velocity V of point P and the flow v of the corresponding point ρ in the image plane is given as follows [9]:

$$v_\rho = f \frac{Z * V - V_Z * P}{Z^2} \tag{2}$$

where: v_ρ is the flow or velocity of point p in image plane, V is the velocity of point P in object space, f is the camera focal length and Z is the height of the imaging sensor. However, this downward vision approach suffers in various scenarios if the drone is flying over a reflective surface, monochrome surface, moving surface, or surfaces with identically repeating patterns [8]. The forward-looking cameras have a higher potential than the downward vision sensor to overcome many of these challenges.

Different visual-based navigation approaches have been proposed, such as the SLAM (Simultaneous Localization and Mapping), MVO (Monocular Visual Odometry), or SVO (Stereo Visual Odometry) [10]. VO can be divided into feature-based methods, direct based methods, and hybrid methods, which is a combination of both [11]. PTAM (Parallel Tracking and Mapping) is a common approach of MVO for UAVs and could be used in augmented reality [12]. PTAM is a feature-based SLAM algorithm that can track many features and map them based on a sequence of images.

[13] employed a video camera to help in navigation, environment mapping, and obstacle avoidance.

III. METHODOLOGY

This section investigates and compares three approaches for heading estimation of small UAVs in indoor scenarios. The first approach uses the downward camera to estimate the heading change using features tracking of the ground in an accumulative fashion. Such solution is commonly used to control UAVs' hovering maneuver and mainly depends on the availability of sufficient feature points in the ground scene, which could be a challenge in many situations such as flying over monochrome floors, highly reflective surfaces or surfaces with repeating identical patterns. These situations are common in indoor scenarios and most common on the ground surfaces.

The second approach simply uses the magnetometer to provide the UAV heading. While magnetometers can offer absolute heading measurements if sufficiently calibrated, their measurements could be significantly affected by the ferromagnetic materials typically found in indoor environments.

The third approach follows the typical visual odometry (VO) technique but with the forward-looking camera and with the removal of potentially moving objects such as pedestrians from the processing steps to enhance the robustness of the orientation change estimation.

Using the forward-looking camera offers scenes with more rich features to be tracked and reduces the chances of challenging conditions such as monochrome surface, repeating patterns compared with the ground scenes.

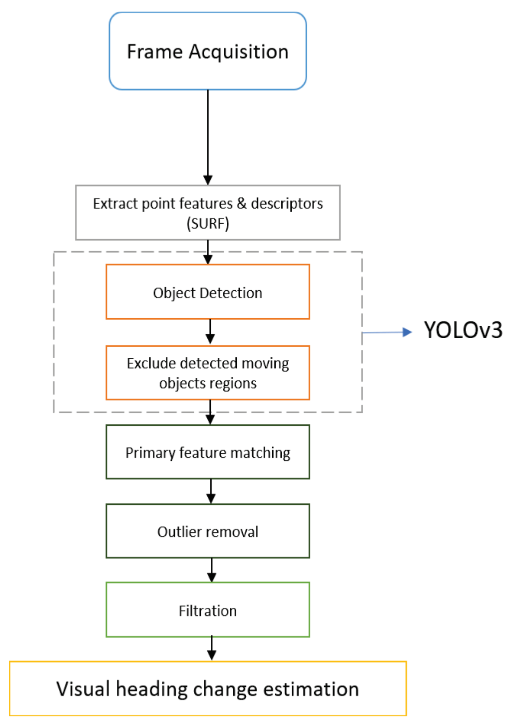

Fig. 1. Algorithm Overview

Figure 1 presents a proposed modified VO algorithm that starts with frame acquisition. The SURF [14] algorithm is then employed to extract the interest points from every frame in the dataset. Matching these interest points between frames helps to estimate the relative orientation between frames given that these points belong to static objects. To avoid the inclusion of moving objects points in the next steps, a machine learning technique for object detection (YOLOv3) [15] is applied to detect moving objects (pedestrians) and exclude them before the matching step. Then the RANSAC (RANdom SAmple Consensus) [16] technique is used to remove the outlier points based on compliance to fundamental matrix model between every two successive frames using the normalized 8-point algorithm [18] using the following steps [17]:

First, both images' coordinates are normalized

$$\hat{x}_i = T x_i \quad (3)$$

$$\hat{x}'_i = T' x'_i \quad (4)$$

where T and T' are the transformations (translation & scaling)

Second, finding a solution F' corresponding to the smallest singular value of matrix A where

$$A\ f = 0 \quad (5)$$

is the set of linear equations built using 8 points

$$\hat{x}\ f\ \hat{x}' = 0 \quad (6)$$

.

Finally, using Singular Value Decomposition (SVD), fundamental matrix F is computed as the closest singular matrix to F' that minimizes Frobenius norm $| F - F' |$.

The estimated fundamental matrix between every two successive frames is used to compute the essential matrix with the help of the internal camera parameters obtained through the camera calibration step. This essential matrix is then decomposed using SVD to estimate the relative camera orientation between these frames [17], where only heading change is considered in this research.

A filtering step takes place to enhance the heading change estimate and to filter undesirable noises by rejecting estimates that exceed a maximum expected derivative value.

The heading estimation is the accumulation of the successive heading change values estimated in the previous steps.

IV. EXPERIMENTAL RESULTS

The performed test is carried out at the CCIT building at the University of Calgary. The traveled distance for the conducted test is around 125 meters. The employed drone in this research is a DJI Mavic mini. The dimensions of this drone unfolded with propellers are (24.5 x 29.0 x 5.5) cm, and its weight is below 250 grams.

Candidate moving objects (persons) are detected and discarded to increase the reliability of the heading estimation.

Figure 2 depicts an example of moving person detection using a deep learning algorithm for object labeling and detection (YOLOv3) [15].

Fig. 3 depicts the matched features between two successive images using the typical VO approach. These matched points could affect the heading change accuracy since there is a moving person in both frames.

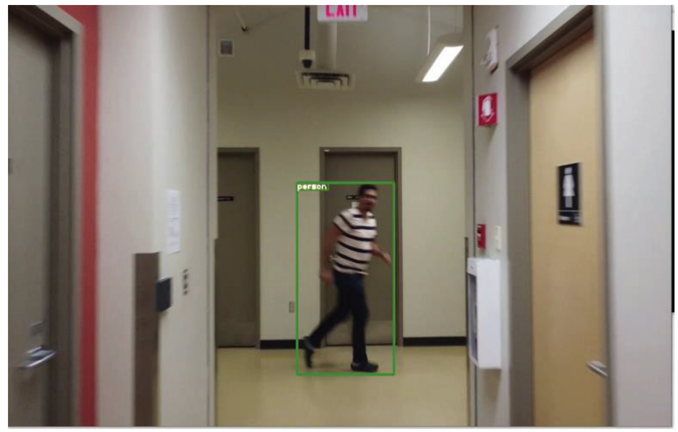

Fig. 2. Detected moving objects such as a person (using YOLOv3)

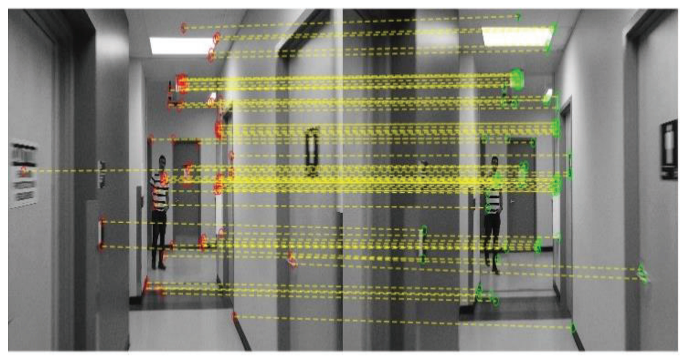

Fig. 3 Matched points between two images (standard matching technique)

The matched points between the two frames, as shown in figure 4 belong only to static objects in the scene after excluding the moving objects.

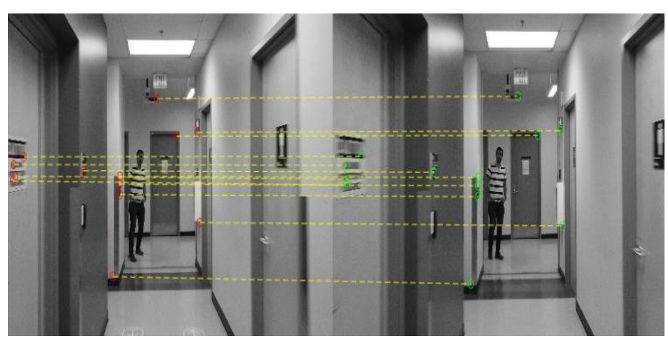

Fig .4 Matched points after moving objects exclusion

Figures (5 and 6) present the estimated change of heading before the filtering step with many spikes and after the filtering step. The filtration step removed outliers at 5.2% of frames which would result in 30.4 degrees of accumulated azimuth error if not properly removed.

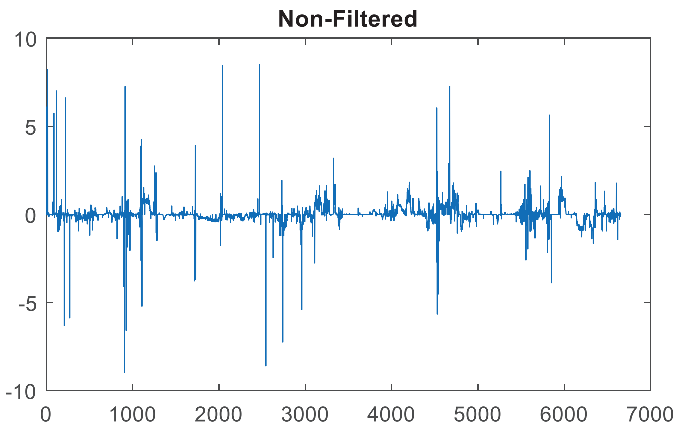

Fig 5. Non filtered change of heading

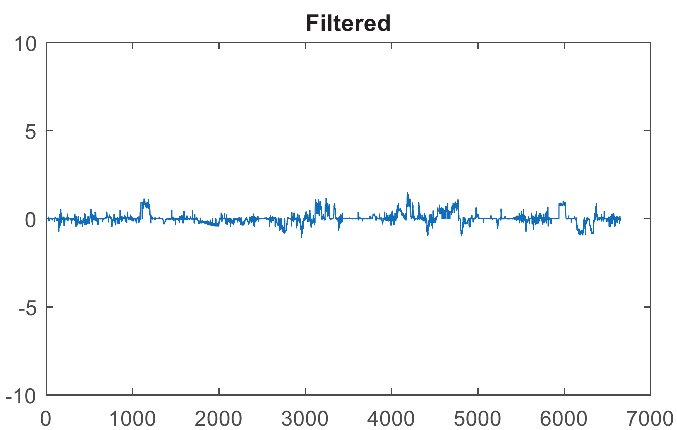

Fig. 6 Filtered change of heading

In order to evaluate the different heading estimation approaches, four trajectories are depicted in figure 7 and overlaid on the map of the test site where the flight started in front of unit 361E near the top of the map and proceeded towards the lower right corridor of the map and returned back through the lower-left corridor. Three trajectories represent the drone path using the magnetometers-driven azimuth, the downward-vision based azimuth, and the proposed heading estimation using the forward camera. The fourth trajectory represents the path using an integrated solution using the downward vision and proposed heading estimation.
While the magnetometer-based trajectory (in red color) aligns with the corridors' directions during most of the test, it suffered from some drift in front of units 301X, 302X (elevator) which could be because of the large metallic doors of these units.
The obtained trajectory based on the proposed heading estimation (in purple color) misses some sharp turns, such as in front of unit 309 as the drone gets closer to the facing wall without enough reliable features to capture.
The obtained trajectory based on the downward camera (in blue color) also drifts in some incidents such as in front of unit 303 where the scene includes different heights downward because of the passed door.
The integrated trajectory represents a compromise between the two involved trajectories and offers the most fitting solution inside the corridors. This result indicates the potential of using the forward camera for heading estimation or as a complement to the downward camera based approach or the magnetometer based approach.

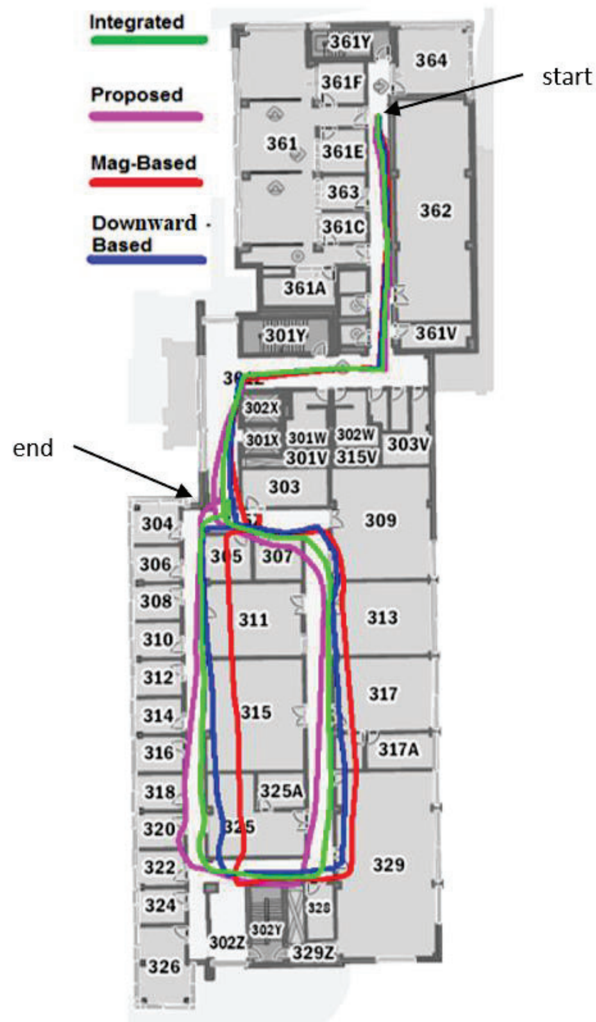

Fig.7 Comparison between 4 trajectories

V. Conclusion

Flying MAVs in an indoor environment is very challenging and typically requires many sensors to help localize the MAV. Cameras are among the primary installed sensors on almost all drones, which encourages its employment for the heading estimation. The proposed approach for VO managed to offer a competitive heading estimation by masking moving persons from camera frames using deep learning. The paper investigates low-cost techniques for heading estimation without depth sensors and compared the estimated drone trajectories based on the investigated approaches. The downward vision sensor has many limitations regarding the ground surfaces, which is common in many indoor scenarios. However, the drone's forward vision sensor could provide a competitive heading estimation that could be implemented solely or integrated with other heading estimation approaches.

Acknowledgment

This research has been conducted under the supervision and funding of Prof. Naser El-Sheimy from NSERC and Canada Research Chairs programs.

References

[1] Transport Canada, "Find your category of drone operation," Government of Canada, 2019. [Online]. Available: https://tc.canada.ca/en/aviation/drone-safety/find-your-category-drone-operation. [Accessed: 02-Nov-2020].

[2] Murphy, Robin R. Disaster robotics. MIT press, 2014. .

[3] M. Faessler, F. Fontana, C. Forster, E. Mueggler, M. Pizzoli, and D. Scaramuzza, "Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle," J. F. Robot., vol. 33, no. 4, pp. 431–450, Jun. 2016, doi: 10.1002/rob.21581.

[4] S. Hrabar, G. S. Sukhatme, P. Corke, K. Usher, and J. Roberts, "Combined optic-flow and stereo-based navigation of urban canyons for a UAV," in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2005, pp. 3309–3316, doi: 10.1109/IROS.2005.1544998.

[5] V. Lippiello, G. Loianno, and B. Siciliano, "MAV indoor navigation based on a closed-form solution for absolute scale velocity estimation using Optical Flow and inertial data," in Proceedings of the IEEE Conference on Decision and Control, 2011, pp. 3566–3571, doi: 10.1109/CDC.2011.6160577.

[6] F. Ruffier and N. Franceschini, "Visually guided micro-aerial vehicle: Automatic take off, terrain following, landing and wind reaction," in Proceedings - IEEE International Conference on Robotics and Automation, 2004, vol. 2004, no. 3, pp. 2339–2346, doi: 10.1109/robot.2004.1307411.

[7] S. Zingg, D. Scaramuzza, S. Weiss, and R. Siegwart, "MAV navigation through indoor corridors using optical flow," in Proceedings - IEEE International Conference on Robotics and Automation, 2010, pp. 3361–3368, doi: 10.1109/ROBOT.2010.5509777.

[8] DJI, "Mavic Mini," 2019.

[9] D. Honegger, L. Meier, P. Tanskanen, and M. Pollefeys, "An open source and open hardware embedded metric optical flow CMOS camera for indoor and outdoor applications," in Proceedings - IEEE International Conference on Robotics and Automation, 2013, pp. 1736–1741, doi: 10.1109/ICRA.2013.6630805.

[10] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in Proceedings - IEEE International Conference on Robotics and Automation, 2014, pp. 15–22, doi: 10.1109/ICRA.2014.6906584.

[11] D. Scaramuzza, "1-point-RANSAC structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," Int. J. Comput. Vis., vol. 95, no. 1, pp. 74–85, 2011, doi: 10.1007/s11263-011-0441-3.

[12] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR, 2007, pp. 225–234, doi: 10.1109/ISMAR.2007.4538852.

[13] S. Ahrens, D. Levine, G. Andrews, and J. P. How, "Vision-based guidance and control of a hovering vehicle in unknown, gps-denied environments," in Proceedings - IEEE International Conference on Robotics and Automation, 2009, pp. 2643–2648, doi: 10.1109/ROBOT.2009.5152680.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, Jun. 2008, doi: 10.1016/j.cviu.2007.09.014.

[15] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv Prepr. arXiv1804.02767, Apr. 2018.

[16] M. A. Fischler and R. C. Bolles, "Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Commun. ACM, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.

[17] R. Hartley and A. Zisserman, Multiple view geometry in computer vision. 2003.

[18] Longuet-Higgins, H. Christopher. "A computer algorithm for reconstructing a scene from two projections." Nature 293.5828 (1981): 133-135.