Real-time On-board Stereo Visual-Inertial Odometry System for Tracking Remote Controlled Vehicle

Botao Hu, Jiajie Zhang, Ji Gu, Scott Macdonald (Amber Garage Inc., California, USA)  
Shuo Yang (DJI Inc., Shenzhen, China)  
Yefang Li, Xiaoyan Yu (Auckland, New Zealand)

I. INTRODUCTION

Using a UAV to track the position and velocity of a remote controlled vehicle is a challenging problem, which have 3 main challenges:

1) **real-time UAV state estimation**: because the UAV is not a stationary observer in the air while tracking, it moves all the times, the UAV itself has to know its current position, rotation and velocity \( p^w, R^w, v^w \) relative the ground frame \( w \);

2) **real-time vehicle state estimation** we need to recover the vehicle’s 3D position, rotation and velocity \( p^c, R^c, v^c \) relative to the UAV’s body frame \( c \) from those 2D computer vision feature points of the remote controlled vehicle in the camera image from a wide angle down-facing camera mounted on the UAV. This problem is also called a classical perspective-n-point (PnP) problem.

3) **UAV follows vehicle** simply hovering the UAV high enough to use a wide angle down-facing camera to cover the whole area may fail to track the vehicle, because the remote controlled vehicle is relatively small in the size 15cmx20cm compared to the tracking area 5mx5m, and camera resolution may not be enough. Hence, we have to hover the UAV in low height and keep following the vehicle. It’s a trade-off between the visual trackability and the UAV’s moving feasibility.

If those three problem are solved, our strategy is very straightforward demonstrated in figure I-C: while the UAV keeps following the vehicle to have it covered in its down-facing camera’s field of view, and then the computer vision system will track the vehicle and give out its state \( p^c, R^c, v^c \) relative the UAV’s body frame \( c \), since the UAV knows its current state \( p^w, R^w, v^w \) relative to the ground frame \( w \), we could retrieve the vehicle’s position and velocity \( p^w, v^w \) relative to the ground frame \( w \) by simply combining those two states together

\[
p^w = p^w + R^w(R^c)^{-1} p^c, v^w = R^w (R^c)^{-1} v^c
\]

Because that those three challenges are tightly mutually coupled. For example, an inaccurate vehicle state estimation might lead the UAV fail to follow the vehicle, so the accuracy, response speed and frequency of the solutions to each challenge are critical to this problem.

We propose our solutions to each of the challenges:

A. **problem 1 real-time UAV state estimation: visual-inertial odometry**

The traditional UAV positioning is using GPS module (e.g. uBlox GPS module), which could only provide 1Hz to 10Hz frame rate. Low frequency of position sensing will be the bottle-neck to track the fast moving vehicle. So an accurate and high frequency onboard localization system is required. So we propose and equipped the UAV with an high-precision visual-inertial odometry (VIO) system to sense the UAVs physical state such as position, rotation and velocity. The system includes visual sensors (5 pairs of mini wide angle stereo cameras (mounted to face four direction (forward, backward, left, right and down-looking) and a high precision and high frequency inertial sensing unit.

Since our system is benefitted from fusing the data from both visual and inerial sensors, the robustness and accuracy would be significantly improved compared to the pure vision-based or pure inertial navigation system. The technology detail of our VIO system is detailed in our Challenge 2 write-up (the real-time UAV state estimation uses the same system that we used in Challenge 2) and published in [1].

Our whole VIO system could provide high precision data streaming of the physical state, such as, position, rotation and attitude of UAV in 200Hz in real-time. Our experiment shows compared to the tracking data of motion capture system, the VIO system could provide the accuracy of 5cm in position error, and 1 deg error in rotation error in 15min hovering tasks from the ground 2m and 10cm in position error and 3 degree error in the task of circling in 5mx5m room.

Additionally, since our system is based on visual and inerial sensors, we dont fully depend on GPS, our system should also work well in a GPS-denied environment. (e.g. close to high building and indoor).

B. **problem 2, real-time vehicle state estimation: visual tracking system**

While the UAV keeps following the vehicle’s movement to have it covered in its down-facing camera’s field of view, we use computer vision technology to detect the vehicle in the image. Since the vehicle has white roof and have fixed and known shape and pattern, we could easily employ contour detection algorithm to extract the 2D contours and features of the remote controlled vehicle from the image. We already knows the size of the vehicle and our camera is calibrated. It’s easy to build an equation system about those feature
and contours’ 3D-to-2D point correspondences. This is a perspective-n-point (PnP) problem, could be easily solved by the algorithm provided in OpenCV library. From the solution, it’s easy to recover the position, velocity, rotation of the vehicle relative the UAV’s body frame $\pi$, which we already know from problem 1.

In our experiment, we first use ARTag pasted on the vehicle to improve the system robustness and prove the concept. Then we finally will use a pure white proof of the vehicle to complete the final challenge.

Because the computer vision system is not always stable, (feature might not lose, contour might not be found at some moment, PnP algorithm might not be always have a solution). We pipe the estimated result from computer vision system into a Unscent Kalman Filter [2] to make the result position and velocity data streaming very smoothly and precisely, and robustly.

C. problem 3, UAV follows vehicle

From problem 1 and problem 2, we already know the position, and velocity of both the UAV and the vehicle. It’s easy to build a PID controller to control position velocity feedback loop to let the UAV to follow the vehicle to keep the vehicle centered in its down-facing camera’s image field of view.

A notable technology detail is that to maintain a reasonable altitude of the UAV is tricky. Since the vehicle’s movement is fast, in order not to lose the vehicle in the camera (the visual trackability), we have to lift high enough. But if the UAV’s hovering too high, the camera resolution is not enough to see the feature of the vehicle, we also lose the track of the vehicle.

We will dynamic adjust the altitude of the UAVs depending on the speed of the vehicle. If the vehicle is moving fast, we will lift up the UAV to prevent it out of the camera. If the vehicle is moving slow, we will lift down the UAV to capture more features to yield more accurate result.

By experiment, we dervied that hovering 2m is a good balanced default altitude to trade off the above problem.

II. Setup

Our base airframe is the recent released M100 (DJI M100) quadrotor, which equipated with a high define 4K video camera (DJI Zenmuse X3) in the front. On top of the quadcopter, we mounted a Nvidia Jetson TK1 as our data processing unit, which provides powerful image processing capability to run our visual inerial odometry (VIO) algorithm to provide high frequency (200Hz) position estimation data stream by taking the input of 5 pairs of the stereo cameras (DJI Guidance, 50Hz) and the inerial sensor (MicroStrain 3DM-GX4-25, in 1000Hz).

Additionally, we mounted a mvBlueFOX global shutter camera with a wide angle lens of 90 degree FOV to track the moving vehicle.

The whole system is under 5kg, battery life could support 20min flight time and could take high quality 4K footage.

Our core technology VIO algorithm is detailed described in [1] and our Challange 2 writeup.

III. Experiment Result.

Experiment setup. We use our UAV to anotomously to track the manually controlled vehicle in motion capture system (accuracy 1mm) tracked 10mx10m indoor environment. We compared our result to the result from motion capture system (1mm accuracy).

Experiment 1. We slowly zigzag the vehicle in the 5mx5m room. the UAV indoor maintain the off-ground distance 2m. Our position estimation error of the UAV is within 5cm in position and 1deg error orientation, the our vehicle estimation is within 5cm and 3 deg orientation error for first 5 min.

Experiment 2. We circle the vehicle in 2m/s speed in the 5mx5m room. UAV will follow indoor at off-ground distance 3-4m. our vehicle position estimation error is within 20cm and 10 degree error in orientation.
IV. COMMERALIZATION POTENTIAL

All hardware parts in our system are off-shelf parts. Nvidia processor, The stereo cameras and inertial measure unit all available in the markets. Our core technology is the software running on the on-board processor, and the VIO algorithm to fuse visual and inertial information together. The software itself is very easy to replant and adapt to any UAV system which provide flight control API, such as both open sourced Pixhawk and DJI Flight Control SDK. So it's very easy to deploy our technology to any brand UAV system.

ACKNOWLEDGMENT

The authors would like to thank DJI for hardware supports.

REFERENCES
