Real-time UAV trajectory generation using feature points matching between video image sequences

Younggi Byun¹, Jeongheon Song² and Dongyeob Han¹

¹Department of Marine and Civil Engineering, Chonnam National University, Yeosu, 59625, Korea
²Calibration Laboratory, High Performance Remote Sensing (Hypersensing), Daejeon, 34133, Korea

E-mail: hozilla@chonnam.ac.kr

Abstract. Unmanned aerial vehicles (UAVs), equipped with navigation systems and video capability, are currently being deployed for intelligence, reconnaissance and surveillance mission. In this paper, we present a systematic approach for the generation of UAV trajectory using a video image matching system based on SURF (Speeded up Robust Feature) and Preemptive RANSAC (Random Sample Consensus). Video image matching to find matching points is one of the most important steps for the accurate generation of UAV trajectory (sequence of poses in 3D space). We used the SURF algorithm to find the matching points between video image sequences, and removed mismatching by using the Preemptive RANSAC which divides all matching points to outliers and inliers. The inliers are only used to determine the epipolar geometry for estimating the relative pose (rotation and translation) between image sequences. Experimental results from simulated video image sequences showed that our approach has a good potential to be applied to the automatic geo-localization of the UAVs system.

1. Introduction

Unmanned aerial vehicles (UAVs) systems have attracted the attention of the research community for their advantage to gain information about surrounding environment [1]. It can be used fruitfully in search and rescue operation, inspection tasks, surveillance, and recognition, so these UAV require high maneuverability and robustness with respect to unexpected system behavior such as GPS system error. Currently, autonomous navigation system for UAV is GPS reliant, but GPS is unavailable in underground, GPS jamming environments, and can also be degrade or denied in certain geographic regions. In order to autonomous navigation of UAVs, there is a need for geo-localization solutions that operate in these condition. As computer vision algorithms mature, many researchers have interested in vision-based navigation systems for overcoming this problem. The DARPA grand challenge has demonstrated the effectiveness of vison sensor in autonomous navigation by analyzing visual information for video cameras mounted on the mobile platform. Effective use of video sensors for navigation has been a goal in ground vehicle robotics for many years. In recent year, many algorithms have been developed, which can be broadly divided into methods using monocular camera and methods using stereo camera [2].

Successful results with a monocular camera over long distance have been obtained in the last decade using both perspective and omnidirectional camera [3]. These approaches can be separated into
methods which either use of feature point matching [4] between consecutive images or feature tracking over a sequence of images [5]. The basic pipeline of feature points-based methods includes a feature point extraction stage, followed by feature matching, epipolar geometry construction using matching points, and finally estimation of relative pose between image frames. Feature tracking-based method require tracking image features (e.g. corners) over a certain number of images.

However, all approaches mentioned above have in common that they only focus on the navigation of ground vehicle robotics or low-altitude UAV. Although successful in other platforms, these methods have not yet been applied to medium-altitude UAV system. In this paper, we describe a real-time method based on feature point matching for deriving medium-altitude UAV trajectory from monocular video sequences. The purpose of this paper is to investigate and evaluate the effectiveness of vision-based navigation system for geo-localization of medium-altitude UAV. Because we cannot acquire video sequence taken by medium-altitude UAV for commercial use, we made a video simulator using Vricon 3D surface model [6] to generate a simulated video sequence for testing. Our method firstly extracts salient features and descriptors from consecutive images in a simulated video sequence. Feature points are then matched in pairs by the similarity of their SURF descriptors. To deal with outliers, a rejection step based on the Preemptive RANSAC is used. The 6 degrees of freedom (6DoF) egomotion is estimated merely from image matching points. We do not restrict the degrees of freedom by using a special (nonholonomic) motion model [7], making our approach widely applicable. The performance of the proposed systems is assessed in terms of both speed and accuracy with respect to reference data.

2. UAV trajectory generation

Generally, video files are made up of thousands of separate images called frames. In order to generate the UAV trajectory, feature points are firstly detected from consecutive images in video file. Detected feature points in previous frame are then matched in current frame. These matching points are used to determine the relative pose of successive images taken by UAV.

2.1. Feature point image matching

In recent years, the feature point detectors and descriptors are employed in many feature-based image matching systems. SURF is one of the most popular feature point detector and descriptor which has been proven to achieve high repeatability and distinctiveness [8]. In this paper, we used the SURF algorithm to find the matching points between image sequences, and removed mismatching by using the Preemptive RANSAC.

![Figure 1. Flow chart of feature points matching.](image-url)
2.1.1. Feature detection. During this step, the image is searched for salient feature points which are likely to match well in other images. Point detectors such as corners or blobs, are important because their position in the image can be measured accurately. In this paper, we used SURF blob detector to extract feature points in images. A blob is an image pattern that differs from its immediate neighbourhood in terms of intensity, colour, and texture. The SURF detector is based on a Hessian matrix with an approximation by using some box filters. The integral image is used to acceleration image convolution in order to speed-up the computations.

2.1.2. Feature descriptor and matching. The region of each detected feature points is converted into a compact descriptor that can be matched against other descriptors. The SURF descriptor is based on an integer approximation of the Haar wavelet responses within a local neighbourhood of each feature point. Feature matching step searches for corresponding features in other images. The general way for matching features between two images is to compare all feature descriptors in the first image to all other feature descriptors in the second image. In this paper, a distance-ratio test was used to find matching points, which accepts the closest match only if the ratio between the closest and the second closest match is smaller than a user-specified threshold.

2.1.3. Preemptive RANSAC. Matched feature points are usually contaminated by outlier. Preemptive RANSAC [3] has been established as an outlier removal method based on breadth-first approach where a fixed number of hypotheses are generated beforehand, and then compared against each other by scoring them in parallel. The hypothesis that shows the highest consensus with the other data is selected as a solution for relative motion estimation. The distribution of features in the image has been found to affect the UAV trajectory results remarkably. In particular, more features provide more stable motion-estimation results than with fewer features, but at the same time, the feature points for hypotheses construction should cover the image as evenly as possible. To do this, the image can be partitioned into a grid, and the sample feature points to compute model hypotheses are selected from each cell randomly.

2.2. Generation of UAV trajectory
The trajectory of UAV can be estimated by using feature points matched across the image sequence. Given accurate matching points in images were used to generate epipolar constraint which was used to estimate relative orientation between images. The relative pose estimation achieved through the essential matrix decomposition.

![Flow chart of relative pose estimation.](image)
2.2.1. Estimating the Essential Matrix. The geometric relations between two images of a calibrated camera are described by the so-called essential matrix. The essential matrix can be computed from matching points using the epipolar constraint. This epipolar constraint is a set of straight lines that are the intersection of the epipolar planes with image plane. In this paper, the normalized 8-point algorithm [9] was used to compute the essential matrix from a set of 8 or more point matches. The great advantage of this algorithm is that it is linear, hence fast and easily implemented. If the 8 point matches are known then the solution of a set of linear equation is involved. With more than 8 points, a linear least squares minimization problem must be solved.

2.2.2. Relative pose estimation. The relative pose (rotation and translation) between two images can directly be extracted from the estimated essential matrix. In general, there are four possible solutions for relative pose for one essential matrix, but, only one geometrically feasible solution. The correct solution can be identified by triangulation of a single point; we put a constraint on the triangulated point that it must be in front of the cameras [10]. The entire trajectory of UAV can be obtained through a concatenation of the relative pose between successive frames.

3. Experimental result and discussion

3.1. Video Data generation

In this paper, we made a simulator using Vricon 3D surface model to generate a simulated video sequence for testing our method because we cannot acquire video sequence taken by medium-altitude UAV for commercial use. Vricon is completely based on commercial satellite imagery and does not require a presence on the ground or height models. It offers high-resolution (0.5m) 3D data whose absolute accuracy is 3 meters in all three dimensions together, SE 90, as proven under operational condition. SE 90 is the abbreviation for Spherical Error 90%, which is a strict measure that combines the traditional LE 90 and CE 90 measures. The simulator is comprised of mission planning simulator and video simulator using 3D data. The mission planning simulator provides the capability to plot the flight path, visually inspect specific events and add visual indicator for test results such as color contoured flight paths. The video simulator generates video sequence along the simulated flight path. In this paper, the mission flew at 4.5 kilometers above ground at Taean-gun in Korea. The video camera was looking downwards and fixed with the mission body. The simulated video sequence is recorded at 30Hz frame rate.

![Mission planning and video simulator](image_url)
3.2. Results and discussion

The simulated video sequence for evaluating the proposed method are suburban area in Taean-gun, South Korea. The total number of frames in a simulated video is 1098, and the experiment is carried out on a PC (CPU: Intel Core-i7 3.6GHz, Memory: 16GB). For the SURF feature matching between image frames, the number of octaves was set to 4, and each octave is subdivided into three scale level. In the SURF-based method, the matched points for which the ratio of the distance of the closest and second-closest were greater than predefined threshold were rejected. If the threshold of the ratio is decreased, the number of matching pairs decrease, but the correct-match rate is increased. We chose this value as 0.6, which is stricter than the standard SURF, because a minimum of 8 matching points is all that is required to construct epipolar geometry from two images. Figure 5(a) is an image of matching result by the SURF method. The crosses show the matching points between two images, and some outliers are detected as shown in Figure 5(a). To remove these outliers, we used the Preemptive RANSAC algorithm in conjunction with random sampling consensus in two images. A number of random samples are taken, each containing 8 point-tracks. The 8-point algorithm is applied to each sample and thus a number of hypotheses, which was set 15 in our experiment, are generated. The hypotheses are scored by a preemptive measure over all the matching pairs in parallel and the hypothesis with the best score is selected. After applying the outlier elimination algorithm, the matching result from the proposed method is shown in Figure 5(b).

![Figure 4](image-url)

**Figure 4.** Video image matching results (100th frame): (a) SURF matching result, (b) result after applying Preemptive RANSAC.

![Figure 5](image-url)

**Figure 5.** UAV trajectory result: (left) simulated flight path, (right) the estimation result of UAV trajectory (red line: True path, green line: Estimated path).
For the quantitative assessment of the proposed method, the correct-match rate, which is how many matching points are correctly matched, was calculated over 200th frame. We considered each matching point pair from the same location on each image as correct matches by visual assessment. The average of correct match rate by the proposed method is 99.8%, and the sum of calculation time for process at each frame is 249.2 milliseconds. Given accurate matching points between image frames were used to determine epipolar geometry for the estimation of UAV trajectory. Figure 5 shows the estimation result of UAV trajectory, and the visual comparison with the true path was performed. At first glance, one can observe that the estimated path showed consistent with actual UAV path. However, upon close inspection of the estimated UAV trajectory result, we can find that the motion error was accumulated along the path.

4. Conclusions

In this paper, we present an approach for estimating the 3D trajectory of medium-altitude UAV based on corresponding image feature points. The proposed approach is based on feature point matching to construct epipolar geometry from two image frame in a video sequence generated from monocular camera. We have demonstrated how the proposed approach worked with simulated video data generated from the Vricon 3D surface model. The experimental results show that the stability and accuracy of 3D trajectory result are relatively satisfactory, but the estimating trajectory from only a monocular image sequence results in the accumulated error along the path. To improve the proposed approach, we are working on a better model of the system, which accounts for the dynamic behavior of the UAV more precisely by using geo-referenced map to reduce the error accumulated in the trajectory.

Acknowledgments

This study was financially supported by an NRF grant (2016R1D1A3B01007236).

References

[1] Esteban C, Rayn H, Chase L and Duke M B 2017 Journal of Imaging vol 3 no 4 pp. 1-14
[2] Yamaguchil K , Kato T and Ninomiya Y 2006 Vechile ego-motion estimation and moving object detection using a monocular camera Proc. Int. Conf. on Pattern Recognition pp.610-613
[3] Nister D 2003 Preemptive RANSAC for live structure and motion estimation Proc. Int. Conf. on Computer Vision vol 1 pp.199-206
[4] Kim Y and Bang H 2017 Vision-based navigation for unnnanned aircraft using ground feature points and terrain elevation data Proc. Int. Conf. on Aerospace Engineering
[5] Dornhege C and Kleiner A 2006 Visual odometry for tracked vehicles Int. Conf. on Safety, Security and Rescue Robotics
[6] Vricon 2015 Vricon satellite 3D data global 3D model with 0.5m resolution www.vircon.com
[7] Bloch A.M , Drakunov S.V and Kinyon M.K 2006 SLAM Journal on Control and Optimization vol 38 no 3 pp. 855-874
[8] Bay H, Ess A, Tuytelaars T and Gool V.L 2008 SURF: Speeded Up Robust Features Computer Vision and Image Understanding(CVIU) vol 110 no 3 pp. 346-359
[9] Longuethiggins H.C 1981 A computer algorithm for reconstructing a scene from two projections Nature vol 293 no 10 pp. 133-135
[10] Nister D 2003 An efficient solution to the five-point relative pose problem Proc. Int. Conf. on Computer Vision and Pattern Recognition pp.195-202