Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings

Volume 15 *Seoul, South Korea* Article 33

2015

An On-Board Visual-Based Attitude Estimation System For Unmanned Aerial Vehicle Mapping

M. Ridhwan Tamjis
School of Civil and Environmental Engineering, UNSW

Samsung Lim
School of Civil and Environmental Engineering, UNSW,

Recommended Citation

Tamjis, M. Ridhwan and Lim, Samsung (2015) "An On-Board Visual-Based Attitude Estimation System For Unmanned Aerial Vehicle Mapping," *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*: Vol. 15 , Article 33.
DOI: https://doi.org/10.7275/R5HM56N9
Available at: https://scholarworks.umass.edu/foss4g/vol15/iss1/33

AT8-03

AN ON-BOARD VISUAL-BASED ATTITUDE ESTIMATION SYSTEM FOR UNMANNED AERIAL VEHICLE MAPPING

M. Ridhwan Tamjis[1] and Samsung Lim[2]

[1]School of Civil and Environmental Engineering, UNSW
Sydney, NSW, 2052, Australia
Email: m.tamjis@student.unsw.edu.au
[2]School of Civil and Environmental Engineering, UNSW
Sydney, NSW, 2052, Australia
Email: s.lim@unsw.edu.au

ABSTRACT

This paper evaluates the performances of several salient feature detectors, namely; Harris detector, Minimum Eigenvalue (MinEig), Scale Invariant Feature Transform (SIFT), Maximally Stable Extremal Region (MSER), Speeded Up Robust Feature (SURF), Features from Accelerated Segment Test (FAST), and Binary Robust Scale Invariant Keypoint (BRISK), in order to assess the suitability in the application of the proposed visual-based attitude estimation system. Throughout the experiment, three main requirements have been investigated which include Time-to-Complete (TTC), detection rate, and matching rate. It was found that SURF fulfills each of the system's requirements. Moreover, it was also found that keypoints detection capabilities affect the processing time, and the clustering patterns in the results may assist in automated inspection of correct and false matching.

1. INTRODUCTION

History of unmanned aerial vehicles (UAVs) can be traced back to as early as 1849 when unmanned balloons loaded with explosives were launched from the Austrian ship to attack the Italian city of Venice (Anonymous, 2012). Since then, the UAV technology is expanding and evolving from a simple balloon to a much sophisticated mechanism. Nowadays, UAVs play important roles in various fields such as engineering, civil, defense, urban planning, recreation, etc. The technology has undergone a series of evolutions in various aspects such as size, material, and control.

The challenges in today's researches on UAVs are to make it more intelligent, robust, and safe during flight, aiming to make an UAV operable on minimal human intrusions (Clough, 2005; Doherty, 2004). One of the approaches is to maximize the vision capabilities equipped in most modern UAVs for attitude estimation. The main idea is to rely on the visual scenes provided by the camera to calculate/estimate attitude. Different types of approaches for visual-based attitude estimation have been proposed in (Garratt & Chahl, 2008; Srinivasan et al., 2004; Srinivasan, 1994).

As an alternative to the conventional sensor-based attitude estimation system, this paper proposes a visual-based attitude estimation system for an UAV and its camera self-calibration, which will use optical flow to measure the egomotion of the on-board camera, and exploit it to estimate the platform's attitude. That is, this study aims to integrate optical flow with a keypoints detector for on-board attitude estimation and camera self-calibration.

This is to minimize the computation load that may be imposed by optical flow. In this paper, the assessment of the possible detectors for the proposed visual-based attitude estimation system is presented.

2. KEYPOINTS DETECTORS

2.1 Visual-based attitude estimation system

The use of visual information for attitude/pose estimation is not new, however, researches are still on-going to discover its potentials. One of the breakthroughs of this approach is on camera self-calibration (Armstrong et al., 1996; Luong & Faugeras, 1997) which has been widely used in modern digital cameras.

The proposed on-board visual-based attitude estimation system is illustrated in Figure 1. The main idea is to use visual information from overlapping images to measure the platform's egomotion, and estimate attitude from the visual motion. Similar approaches have been studied by Dusha et al. (2007) and Srinivasan (1994). Optical flow computation could be expensive depending on the approach (Fleet & Weiss, 2006). For that reason, this research aims to reduce the computation load at the start of the process by limiting the images to regions of upmost important. This requires an integration of optical flow with salient feature detection and matching.

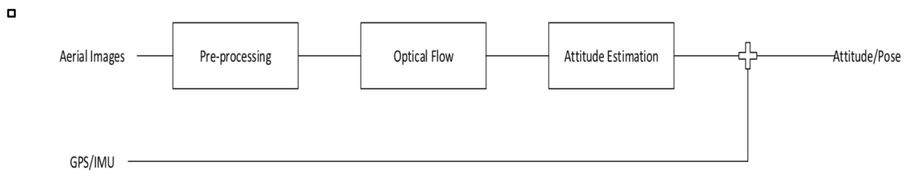

Figure 1. Framework of the proposed system.

2.2 Keypoints Detectors

Several types of detectors are investigated and evaluated in this experiment, namely; Harris detector (Mikolajczyk & Schmid, 2001), Minimum Eigenvalue (MinEig) (Shi & Tomasi, 1994), Scale Invariant Feature Transform (SIFT) (Lowe, 1999), Maximally Stable Extremal Region (MSER) (Matas et al., 2004), Speeded Up Robust Feature (SURF) (Bay et al., 2006), Features From Accelerated Segment Test (FAST) (Rosten & Drummond, 2006) and Binary Robust Scale Invariant Keypoint (BRISK) (Leutenegger et al., 2011). Each of the detectors possesses a different computation algorithm as well as capabilities.

3. APPROACH

A set of aerial images acquired from an UAV is used in the evaluation. The image set consists of 249 overlapping images, taken above Loftus Oval, New South Wales, Australia. The aerial scene shows open yards, vegetation, roads and partial suburbs. The images also show pose variations and distortions which occurred during flight. The selection of detectors for this experiment is based on those attributes, as it was proven that the detectors are invariant to such transformations and distortions in literature.

3.1 Keypoints detection and matching

The main purpose of this experiment is to evaluate the detection rate and matching rate of each nominated keypoints detector and descriptor. For a fair judgement, a standard process framework is required. Figure 2 illustrates the standard detection and matching process that have been used throughout the experiment.

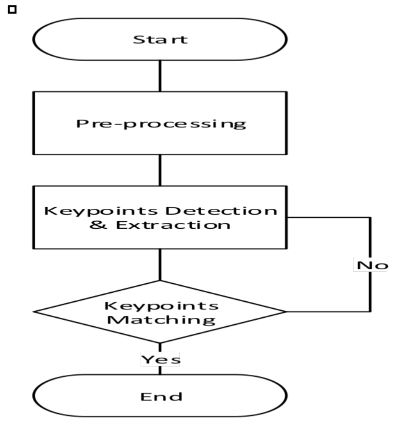

Figure 2 Standard framework for keypoints detection and matching

3.1.1 Image pre-processing

To minimize the complexity and completion time, images are down-sampled to 20% of the original size, an optimal size which does not compromise image features. Furthermore, images are converted to grayscale from its original Red-Green-Blue (RGB) color space to fulfill the processing requirements of keypoints detectors.

3.1.2. Keypoints detection and extraction

The purpose of this process is to detect and extract salient feature points in the image using keypoints detectors. Detection and extraction of the keypoints are conducted in every two consecutive overlapping images in the image sequence. Based on the type of detectors, keypoints are projected to the image, and the locations as well as the descriptors are extracted for the latter matching process.

3.1.3 Keypoints matching

It is the research concern that each detector will work differently in different matching metric. To investigate this issue, two different metrics have been used in the experiment, namely; Sum of Squared Differences (SSD) given in Equation (1), and Sum of Absolute Differences (SAD) given in Equation (2). For two images *f(x,y)* and *g(x,y)*;

$$SSD(d_1,d_2)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(f(x+i,y+j)-g(x+i-d_1,y+j-d_2))^2 \qquad (1)$$

$$SAD(d_1,d_2)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}|(f(x+i,y+j)-g(x+i-d_1,y+j-d_2))| \qquad (2)$$

3.2 EVALUATION CRITERIA

3.2.1. Detection rate

In this experiment, the number of keypoints that can be detected by each detector is investigated. This criterion is chosen to evaluate the capability of each detector to provide a sufficient number of keypoints for the matching process. It is assumed that the probability of obtaining correct matches increases when a higher number of detected keypoints are available, making the matching process robust. Such an advantage has been shown in the SIFT framework, although it will also increase computation time. Therefore, this factor should be considered for on-board applications.

3.2.2. Time-to-Complete

Time-to-Complete (TTC) is measured from the start of keypoints detection to the end of matching, in which matching pairs have been identified. In this evaluation, it is important to select the detector with minimal processing time, e.g. the fastest detector. There are other requirements that are significantly important for the proposed system, however, the highest weight is assigned to TTC.

3.2.3. Match rate

Matching rates of keypoints detectors are determined by the number of correctly matched keypoints over the number of overall matched keypoints. To ensure that the results are statistically significant, an optimal sample size of 151 is chosen based on the sample size calculation, employing 95% confidence interval.

4. EVALUATION OF PERFORMANCE

4.1 Detection capabilities

The experiment results show that BRISK has failed to locate keypoints in at least 60% of the total overlapping images, which is followed by FAST but with a smaller percentage. This is probably due to the capabilities of the detectors to locate keypoints in overlapping images. Furthermore, it is found that there is a relationship between the number of detected keypoints and the number of failed matches, in which the number of failed matches lessens with an increase of the number of keypoints. The same pattern can be observed in the number of matched keypoints.

Table 1. Mean TTC

Keypoints Detector	Mean TTC (seconds)	
	SSD	SAD
BRISK	0.259	0.133
SURF	0.261	0.151
FAST	0.241	0.120
SIFT	1.441	1.410
MSER	0.542	0.403
Harris	0.441	0.324
MinEig	0.811	0.636

4.2 TTC Comparison

The processing time for the detectors to complete the full cycle of the algorithm has been measured in this experiment and the results are shown in Table 1. From Table 1, it can be seen that the processing time varies for each detector while SIFT took the most time to complete the process. This has been expected because of the number of generated keypoints. However, even though MinEig generates a much larger number of keypoints than SIFT, the binary nature of the detector has made the processing time much faster.

4.3 Matching capabilities

In this experiment, it is very difficult to compute the number of correct and false matches in SIFT-processed images due to the density of the matching keypoints, as the lines connecting the points which have been used as an aid for visual inspection rendered the matching points in corresponding images. To tackle the rendering constraints, a different kind of assessment is formulated for SIFT which does not deviate from the comparison goals. For SIFT, correct matches dominate the lines compared to false matches, thus, led to a conclusion that the correct-match percentage for SIFT is particularly high. The other keypoints detectors are assessed straightforwardly, and the results for all detectors (except for SIFT) are tabulated in Table 2.

Table 2. Correct-match results for each keypoints detectors except for SIFT

Keypoints Detector	Correct-Match Percentage (%)	
	SSD	SAD
BRISK	88	82
SURF	61	96
FAST	82	86
MSER	48	94
Harris	91	91
MinEig	93	94

4.4 Assessment and findings

In order to find the best detector for the proposed system, an occurrence-based (best-fit) assessment has been conducted. The results are divided into five selection criteria, namely, (1) TTC, (2) keypoints density, (3) match density, (4) match failure, and (5) percentage of correct matches. Four best detectors are ranked based on the performance in each criterion. The main findings are described below.

4.4.1 Optimal detector

From the assessment, SIFT is ranked first in the performance comparison as it generates the highest number in matched keypoints and correct-match, as well as minimal match failure. However, SIFT is slow in processing, which is a critical criterion in the assessment. The same goes to MinEig. If the processing time is not critical in the assessment, this research highly suggests SIFT for its robustness. In overall, SURF has the optimal performance required by the proposed system, given that SAD metric is employed.

4.4.2 SSD vs SAD

The experiment results have shown that matching metric works differently for different types of detectors. The results of BRISK, FAST, Harris, and MinEig show that the matching metric does not directly affect these binary-type detectors in terms of detection and matching capabilities. On the contrary, integer-based detectors such as SIFT, SURF and MSER are directly affected by it. The SAD metric deteriorates the matching capabilities of the detector, however, increases the number of correct-matches. In terms of computation time, algorithms ran faster when SAD is employed.

4.4.3 Keypoint density vs TTC

It is also found that the detection capabilities affect TTC, in which, a higher number of detected keypoints leads to slower computation time. This is, however, subject to the complexity of the detector's framework.

4.4.4 Clustering pattern

From visual inspection, clustering patterns are detected on the matching points, which may assist in the automated inspection of correct and false matching. The formation of a cluster is dependent on the matching keypoints. Investigation has shown that correct matches can be grouped into one cluster based on the attribute used in the measurement, and made up the large portion of the matched keypoints.

5. CONCLUDING REMARKS

Performances of various keypoints detectors have been evaluated in terms of detection rate, TTC and matching rate. A set of 249 aerial images taken using a fixed-wing UAV's camera have been evaluated, which represents actual flight conditions, translations, rotations, illumination changes, anomaly, and perturbations. Assessments are conducted based on the chosen criteria, which aims to fulfill the UAV's on-board applications requirements.

The assessment results show that the best detector candidate to be integrated to the proposed system is SURF, given that the SAD metric is used to measure similarity between keypoints. It was found that the time taken for SURF to complete the full cycle of the algorithm is relatively small. SURF is also able to provide a sufficient number of salient feature points in each detection without sacrificing the computation time. Additionally, SURF is able to provide a high percentage of correctly matched keypoints, which fulfills the requirements of the proposed visual-based attitude estimation system and camera self-calibration.

REFERENCES

Anonymous. (2012, 15 November 2012). UAV Evolution - How Natural Selection Directed the Drone Revolution. *army-technology.com.* Retrieved 23 April 2015, 2015, from http://www.army-technology.com/features/featureuav-evolution-natural-selection-drone-revolution/

Armstrong, M., Zisserman, A., & Hartley, R. (1996). Self-calibration from image triplets *Computer Vision—ECCV'96* (pp. 1-16): Springer.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features *Computer vision–ECCV 2006* (pp. 404-417): Springer.

Clough, B. T. (2005). Unmanned aerial vehicles: autonomous control challenges, a researcher's perspective. *Journal of Aerospace Computing, Information, and Communication, 2*(8), 327-347.

Doherty, P. (2004). *Advanced Research with Autonomous Unmanned Aerial Vehicles.* Paper presented at the KR.

Dusha, D., Boles, W., & Walker, R. (2007). *Attitude estimation for a fixed-wing aircraft using horizon detection and optical flow.* Paper presented at the Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on.

Fleet, D., & Weiss, Y. (2006). Optical flow estimation *Handbook of Mathematical Models in Computer Vision* (pp. 237-257): Springer.

Garratt, M. A., & Chahl, J. S. (2008). Vision ‐ based terrain following for an unmanned rotorcraft. *Journal of Field Robotics, 25*(4 ‐ 5), 284-301.

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). *BRISK: Binary robust invariant scalable keypoints.* Paper presented at the Computer Vision (ICCV), 2011 IEEE International Conference on.

Lowe, D. G. (1999). *Object recognition from local scale-invariant features.* Paper presented at the Computer vision, 1999. The proceedings of the seventh IEEE international conference on.

Luong, Q.-T., & Faugeras, O. D. (1997). Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision, 22*(3), 261-289.

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing, 22*(10), 761-767.

Mikolajczyk, K., & Schmid, C. (2001). *Indexing based on scale invariant interest points.* Paper presented at the Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on.

Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection *Computer Vision–ECCV 2006* (pp. 430-443): Springer.

Shi, J., & Tomasi, C. (1994). *Good features to track.* Paper presented at the Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on.

Srinivasan, M., Zhang, S., Chahl, J., Stange, G., & Garratt, M. (2004). An overview of insect-inspired guidance for application in ground and airborne platforms. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 218*(6), 375-388.

Srinivasan, M. V. (1994). An image-interpolation technique for the computation of optic flow and egomotion. *Biological Cybernetics, 71*(5), 401-415.