

## EXTRACTING ROAD FEATURES FROM AERIAL VIDEOS OF SMALL UNMANNED AERIAL VEHICLES

Deepak Rajamohan\*, K.S. Rajan

Lab for Spatial Informatics, International Institute of Information Technology, Hyderabad, India

\*deepak.rajamohan@research., #rajan@iiit.ac.in

Commission I, WG I/4

**KEY WORDS:** Small UAV remote sensing, traffic flow, geo-referenced data collection, road features extraction.

### ABSTRACT:

With major aerospace companies showing interest in certifying UAV systems for civilian airspace, their use in commercial remote sensing applications like traffic monitoring, map refinement, agricultural data collection, etc., are on the rise. But ambitious requirements like real-time geo-referencing of data, support for multiple sensor angle-of-views, smaller UAV size and cheaper investment cost have lead to challenges in platform stability, sensor noise reduction and increased onboard processing. Especially in small UAVs the geo-referencing of data collected is only as good as the quality of their localization sensors. This drives a need for developing methods that pickup spatial features from the captured video/image and aid in geo-referencing. This paper presents one such method to identify road segments and intersections based on traffic flow and compares well with the accuracy of manual observation. Two test video datasets, one each from moving and stationary platforms were used. The results obtained show a promising average percentage difference of 7.01% and 2.48% for the road segment extraction process using moving and stationary platform respectively. For the intersection identification process, the moving platform shows an accuracy of 75% where as the stationary platform data reaches an accuracy of 100%.

## 1. INTRODUCTION

### 1.1 Motivation

Flying small UAVs is becoming more than a hobby and is finding numerous applications in remote sensing from supplementing data gathering to map refinement. Availability of video and very high resolution images from these platforms open up new possibilities but as well necessitates the use of a range of new video/image processing techniques instead of traditional remote sensing image processing. Also data obtained from a small UAV is mostly oblique and noisy due to its unstable platform. Hence an automatic geo-referencing of the collected data like traffic flow needs additional methods, rather than utilizing only the noisy GPS and attitude sensors.

### 1.2 Background

Currently the video from unmanned aerial vehicles are primarily used in manual monitoring applications. But they are also rarely used in a variety of data collection application as well, one of the example applications being automatic collection of traffic information. Since such information would be more useful when geo-tagged, accurate and automatic geo-registration techniques are desired. One way to achieve this is to use the localization sensors like attitude indicator and the GPS. But in small unmanned aerial platforms, due to in-stability of the platform and sensor noise the automatic geo-referencing of the collected data is not possible. (Figure 1 shows an example scenario where the road centreline map data is overlapped with a real scene using the localization sensor inputs). Hence we need additional techniques to register the real world scene with the map database and then update the collected data on to the geo-database. The idea here is to collect information from the

existing sensors like the video camera and process them in such a way that they can be used in aiding accurate geo-referencing.



Figure 1 : Video frame and map overlay

Road networks are a good source of spatial information and can be used for aiding geo-referencing as the network pattern can be exploited for performing local localization. This local information can be used to refine the approximate global location given by the localization sensors. For this purpose the complete road network need not be extracted but the sparsely located features like road intersections and road segments would be sufficient. These intersections and road segments can aid in correcting the geo-referencing errors by matching these features with the map database. This method works well in urban regions where there is a good amount of traffic flow on the roads.

### 1.3 Related Work

This paper presents a method that uses traffic flow extracted from aerial videos to identify the features of interest - the road segments and intersections. Existing work **Error! Reference source not found.** **Error! Reference source not found.** uses

motion cues in aerial videos to yield active contours so that roads of different shapes can be extracted. The paper **Error! Reference source not found.** mainly focuses on complete road extraction to largely update an existing map, while the focus of this work is to extract additional, though and limited information that can supplement the noisy localization sensors in improving the geo-referencing accuracy. Most of the prior works **Error! Reference source not found.** **Error! Reference source not found.** **Error! Reference source not found.** deal with the road extraction using static aerial or satellite imagery. Since small aerial platforms in consideration are capable of providing very high resolution imagery, the image only techniques do not produce good results. Where as techniques exploiting the spatio-temporal information is a good alternative and is also a less explored research area. With the ever advancing processing power and efficient image processing algorithms that are being developed these techniques will be able to solve complex real time problems.

## 2. METHOD OVERVIEW

### 2.1 Process Block Diagram

The paper focuses on the extraction of the road information – road segment data and road intersection identification; from the video data. The overall process flow is shown in Figure 2 below, and is explained in the following sub-sections.

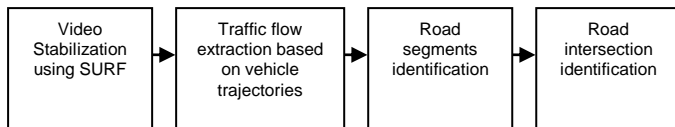


Figure 2 : Road Information Extraction Overview

### 2.2 Image Stabilization

In order to extract foreground objects from video of a moving camera, the video frames have to be stabilized. Successive video frame images are registered using SURF features discussed in **Error! Reference source not found.** to form a mosaic. For a moving platform, continuous mosaic formation leads to distorted images as the platform moves away from the initial position. Hence the mosaic process is restarted after N successive frames. The latest registered frame in the mosaic-video appears stabilized and hence shows moving objects. The successive frames are then subtracted to form the difference image sequence. This image sequence is then processed to find the foreground objects. Large foreground objects extracted are assumed to be the vehicles moving on the road.

### 2.3 Traffic Flow Extraction

Traffic flow is the cumulative direction of vehicles moving on a road. The method used to determine the traffic flow is by tracking the moving objects on the road. Centroids of the foreground objects extracted in section 2.2 are tracked using Lucas-Kanade ‘optical-flow tracking’ algorithm and their trajectory across successive video frames is extracted. Brightness consistency and small movements of objects are assumed to enable tracking by Lucas-Kanade algorithm. When the objects are continuously tracked, the algorithm results in distorted trajectories especially when the moving objects pass through occlusions. This is because the occlusion causes the image sequence to violate the basic assumptions made for tracking and causes the failure of the classic equation (1) of 2D

motion at a pixel discussed in **Error! Reference source not found.**].

$$I_x u + I_y v + I_t = 0 \quad (1)$$

$I_x, I_y$  - Image gradient in x and y direction,  $u, v$  - x and y component of object velocity,  $I_t$  - derivative of image over time.

The tracking is lost and the tracked image points are likely to be picked up by other moving objects passing through the same position and hence resulting in distorted trajectories. The tracked points are hence cleared after a pre-defined number of frames and new centroids are fed into the tracking algorithm. This reduces tracking errors due to occlusion.

In order to detect the traffic flow direction on each road, the traffic flow at the road junctions and corners need to be neglected. So the trajectories that have higher than an empirical RMS of deviation value from their line-of-best-fit, (see Figure 3) are rejected and only the linear ones are selected for further processing. RMS of deviation value is calculated by equation (2).

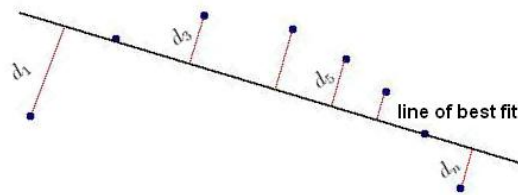


Figure 3 : RMS of Deviation from Line-Of-Best-Fit

$$RMS = \sqrt{\frac{d_1^2 + d_2^2 + \dots + d_n^2}{n}} \quad (2)$$

Now the trajectories of different road segments need to be grouped together to provide for a unique road segment. Slope and centre point of the line-of-best-fit for each trajectory is found (Figure 5) so that a three dimensional vector space (Figure 4) constituted by slope on one axis and centre point x and y image coordinates on the other two axes.

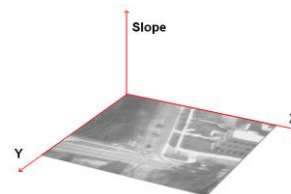


Figure 4 : (slope, x, y) Vector Space

Using the slope and location of the trajectory, its parent road segment is identified. This is done using a Minimum Spanning Tree on the vector space. The MST is then clustered (Figure 5) based on a defined threshold, which further depends on the camera angle-of-view.



Figure 5 : Trajectories of Moving Objects & MST Clustering

Clustering could result in grouping of trajectories belonging to different roads even if the roads are at right angle to each other. Trajectories of two roads tend to form a single cluster if there is only a gradual change in the slope of the trajectories (Figure 6).



Figure 6 : Trajectories in the Road Junctions Tend to Form Incorrect Clusters

This scenario is handled by splitting the curved clusters into two or multiple linear clusters based on the RMS of deviation values from their line-of-best-fit, in a similar approach as that for the curved trajectories. The RMS deviation is recalculated successively for each addition of a trajectory to the cluster and used to split the cluster appropriately when the RMS value exceeds a threshold.

#### 2.4 Finding Road Segments and Intersections

Each cluster gives the traffic-flow on a particular segment of road since it shows the direction of the traffic flow and hence the direction of the road. The cluster is a representation of road in itself. Road intersections are points where at least two road segments with more than an acceptable degree of slope difference meet. These clusters are replaced by their lines of best fit and then extended with constraints like cluster proximity, angle to their line-of-best-fit. This helps in identifying the intersections (Figure 7). Multiple intersections

Cluster No	Angle (degree)		Abs Diff	% Diff
	Algorithm	Manual		
1	125.429667	128.4108683	2.981201333	2.348877282
2	47.53184	47.51475533	0.017084667	0.035950087
3	2.80637	2.650245667	0.156124333	5.72238702
4	167.988505	171.119344	3.130839	1.846515207
Average % Difference				2.488432399

are formed with one from each pair of clusters. Based on proximity, a single mean intersection point is found.

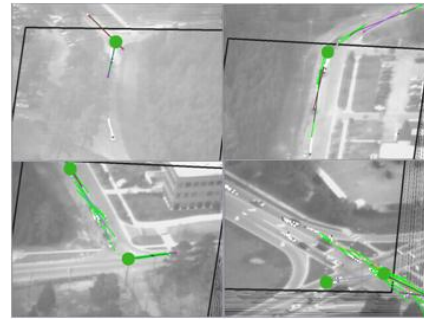


Figure 7 : Road Intersection Identification

### 3. RESULTS

Since we depend on real time traffic for the extraction of road segments and intersections, the performance of the extraction algorithm depends on how well the irregularities in traffic are handled. In this paper, a comparison between the video data collected by a stationary versus a moving platform has been attempted. How accurate the extracted road segments matches with the real road decides the performance of algorithm. Though it is difficult to come up with a robust, consistent measure of accuracy, we compare it with the accuracy of an expert based manual extraction process. So the accuracy of the road segment extraction was measured by finding the difference between the slope of the identified traffic flow cluster and the slope of the road found manually by drawing a line along the road in the image. The difference between these angles divided by their average provides the percentage difference. The results for videos from both stationary and moving platforms are

Cluster No	Angle (degree)		Abs Diff	% Diff
	Algorithm	Manual		
1	150.780762	153.634583	2.853821	1.874952132
2	75.57922	76.255086	0.675866	0.890267842
3	162.897257	160.534139	2.363118	1.461279288
4	-22.963773	-22.619865	0.343908	1.508909842
5	-164.57784	-160.232829	4.345011	2.675411503
6	-100.234802	-100.69222	0.457418	0.455307599
7	67.520569	67.041946	0.478623	0.711376419
8	-148.822052	-146.398773	2.423279	1.641672128
9	174.522624	164.787415	9.735209	5.738238119
10	139.969747	139.969747	0	0
11	-39.17366	-38.59256	0.5811	1.494479222
12	113.498566	109.698629	3.799937	3.405004261
13	167.374166	171.093262	3.719096	2.197609396
14	159.848246	156.997406	2.85084	1.799513411
15	57.0062	32.763891	24.24231	54.00976813
16	-106.858398	-148.099833	41.24144	32.35152271
Average % Difference				7.013457

shown in Table 1.

(a) Moving Platform Dataset results

(b) Stationary Platform Dataset Results

Table 1: Results of Road Segment Extraction

The result shows that the extraction of road segments is more accurate in case of a stationary camera compared to a moving one. Intersection identification accuracy is measured by tabulating the results in Table 2 using the equation (3). True

positives are the count of intersections that matches a visual identification where as the false negatives are intersections that do not match with the visual identification. False negatives and true negatives were counted as zero as they can not be measured in this scenario. The algorithm was able to identify the single road intersection that was present in the chosen stationary camera video sequence. In the moving camera video sequence, out of six intersection points all six were identified. Two erroneous points were also identified as intersections.

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100 \quad (3)$$

	True	False		True	False
<b>Positive</b>	1	0	<b>Positive</b>	6	2
<b>Negative</b>	0	0	<b>Negative</b>	0	0

(a) Stationary Platform Results

(b) Moving Platform Results

Dataset	Accuracy %
Stationary Platform	100
Moving Platform	75

(b) Results in Percentage

Table 2: Results of Road Intersection Extraction

#### 4. CONCLUSION

Results suggest that this method of road segment extraction and road intersection identification is quite promising for videos from UAVs that can hold their position in air. Reduction of accuracy in moving platforms was due to the building and other occlusions. Addressing this problem by depth perception is planned for future work. But an importation continuation of this work is to use the identified road segments and intersections in improving the geo-referencing accuracy. Also other practical issues like the execution time of the image stabilization algorithm need to be improved for any real-time application. The empirical values in this work have been calculated by considering the UAV camera to be at an angle of 45 degrees. Hence the future work will also attempt to make the algorithm more flexible in this regard. Overall, the above procedural method is a new but simple way of processing the spatio-temporal information available on the aerial videos keeping in mind the application of geo-referencing.

#### 5. REFERENCE

B.D. Lucas, T. Kanade., 1981 An Image Registration Technique with an Application to Stereo Vision, in Proceedings of Image Understanding Workshop, pp. 121-130

Berthold K.P. Horn, Brain G. Schunck., 1981 Determining Optical Flow, Artificial Intelligence 17 (1981) 185-203

David A. Jurgens., 2004. Road Extraction from Aerial Videos using Active contours and Motion Cues.

Junhee Youn, James S. Bethel, Edward M. Mikhail, and Changno Lee., Extracting Urban Road Networks from High-

resolution True Orthoimage and Lidar, Photogrammetric Engineering & Remote Sensing Vol. 74, No. 2, February 2008, pp. 227-237

Luo Juan, Oubong Gwun., A Comparison of SIFT, PCA-SIFT and SURF, International Journal of Image Processing (IJIP) Volume(3), Issue(4).

Ong Hen Ching., 2010 Optical Flow-Based Vehicle Detection and Tracking.

S.Hinz, A.Baumgartner, Automatic Extraction of Urban Road Networks from Multi-view Aerial Imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 58:83-98, 2003.

Reinhold Huber, Konrad Lang., Road Extraction from High Resolution Air-borne SAR using Operator Fusion. Proceedings of International Geoscience and Remote Sensing Symposium, July 2001.