Multiplatform Stereoscopic 3D Terrain Mapping for UAV Localization
Auburn REU on SMART UAVs 2018
CSSE18-04

Marchio, Jacob
University of Maine
ejacob.marchio@maine.edu

Kinzel, Charles
Cal State San Bernardino
charleskinzel@yahoo.com

Chapman, Richard
Auburn University
chapmro@auburn.edu

Biaz, Saad
Auburn University
biazsaa@auburn.edu

July 18, 2018

Abstract

This paper presents a study of UAV homing in a GPS denied environment. More specifically it focuses on a system of drones, in this case two, which localize when GPS becomes degraded or non-existent. The results show that in general, an initial overlap of 78% was too little to generate disparity using the Semi-Global Block Matching algorithm, while an initial overlap of 92% was very usable. For analysis, 8 sets of 5 scenes were used to generate disparities. Using a Normalized Cross-Correlation Matching algorithm, the depth map was found on the global depth map and the location and orientation of the UAVs are calculated in every instance that a successful disparity map was generated.

1 Introduction

Flight has changed greatly in the last several decades. Aircraft increasingly have required less human attention, and much flight happens without a pilot in the cockpit; indeed, with vehicles small enough to have no cockpit. These Unmanned Aerial Vehicles (UAVs) have great advantages, but also pose great challenges. For example, navigation is one of the major challenges for UAVs. Because there is no direct human pilot on board, there must be extensive sensing and localizing in order to control and guide the craft. Human pilots have the distinct ability to observe, adjust, communicate with ground and other aircraft, and make situational judgments on the fly. On the other hand, UAVs must rely on on-board or external computations and communications. The autonomous flight of UAV in GPS denied environments has been a hot spot for research and continues to gain attention in many areas of science and technology. The applications are endless and range from small business logistics to planet exploration and everything in between.

The search for an optimal navigation system for the case when GPS signals are degraded or non existent is currently under research. UAV homing without the aid of GPS starts with the vehicle localizing itself. It must have the ability to determine its location with respect to a starting point, a destination, or a map. Inertial navigation systems (INS) can provide a starting point, but due to the nature of the system accumulated errors overwhelm the navigation. In many cases researchers accomplished the process of localization with local map generation using various technologies. Whether using LiDAR, range finders, stereo vision, Google
Maps or some other technology, a map is generated of the surrounding area. Once a UAV can predict its location with some level of accuracy it can begin the process of navigating to its destination. Scientists have tackled this problem with feature extraction algorithms and other navigation systems like SLAM and optic flow.

This paper will examine a new method of local map generation to localize a GPS-denied UAV. Using two UAVs (or more, this model can be extended to \( n \) UAVs), each with a nadir oriented camera, a capture of the terrain below can be reconstructed using stereo vision methods—specifically Semi Global Block Matching (SGBM)—to form a 3-dimensional depth map of the terrain. This depth map will be compared to a locally stored terrain map of the area. This stored terrain map could be built from widely available topographical maps and loaded to the UAV before flight in the area. By comparing the 3D capture of the topography beneath the UAVs to the stored terrain map, the location of the UAVs can be determined. This paper will look at a simulation of this using virtually generated terrain and cameras to generate disparity maps. This disparity map will then be used as a height map to compare to a much broader map of the region using template matching, which will localize the UAVs.

2 Prior Work

Various methods have been implemented to solve the problem of UAV homing in a GPS denied environment. The CS department at the University of Wales conducted research in a simulated environment of Mars. A laser range finder module was used to generate a local map and feature extraction algorithms were used to localize and navigate to a pre-determined destination [10]. This project shares several aspects with the research topic of this paper, namely that a 3D map is rendered and then compared with an existing global map and the result is used for localization. The project is different in several key aspects that this paper implements, mainly, the experiment contained herein implements the use of stereo vision on multiple UAVs to localize. Research has also been done on wide-baseline stereo imaging for localization and navigation. One experiment conducted by Olson and Abi-Rached, was done with Martian applications using this technology. A single camera was used to take multiple images that were then rectified to produce depth maps which could then be used by the robot [1]. This research is important to this paper as it uses a wide base line stereo vision to render a terrain map that is used to localize itself. The methods differ in that the photos are taken by one ground robot and what is considered a wide base line is 20 centimeters to 2 meters. The base line considered for the experiments in this paper are on the order of 100 meters. Another study investigated the possibility of a decentralized flocking system of UAVs to localize and find a target. Each drone in the system was equipped with a front facing camera and downward facing LiDAR. The cameras were used to track their leader (or the target in the case of the lead UAV), while the LiDAR was used to collect the data. Each were programmed using a SLAM system to explore and learn the environment [4]. This research relates to this paper by its use of multiple drones in the flocking system, but like the previous work mentioned, there is no use of a wide baseline stereoscopic terrain mapping.

Concerning 3 dimensional reconstruction of terrain, Researchers have previously used UAVs to model terrain in the context of geoscience using structure from motion [5]. Using a combined structure from motion and multi-view stereo technique, M. James and S. Robson measured relative precision of greater than 1:1000 (1 cm precision at 10 meter elevation) with a UAV mounted DSLR camera, which is close to the theoretical estimated by stereo photogrammetry for their setup [9]. One way to navigate in 3D environments is to use a "voxel" concept; Dryanovski, Morris, and Xiao introduced a method for mapping 3D environments called Multi-Volume Occupancy Grids. This method was open source, could be updated in real time, used comparatively less memory and was designed for navigating small UAVs [6]. This is similar to a method presented by A. Hornung, et al, that mapped 3D environments using octrees to create an "OctoMap", a 3D representation of the environment that also reduced memory requirements and allowed 3D robotic navigation [3].

Terrain-aided navigation was implemented by Jiang Wu, Wenkai Fei, and Qianru Li in a system using a laser scanning sensor to gather depth information about terrain. Their method used GPS and INS (Inertial
Navigation System) to help construct the terrain data from the scan, which was used to build a 3D dataset of the terrain with which to navigate by [7]. In a paper authored by David Gallup, et al., stereo depth resolution with variable baselines and camera resolution was examined. The goal was to create an algorithm that, using changes in baseline and resolution, gave a constant depth resolution, regardless of distance from the cameras (depth resolution would normally decrease by a quadratic power from a fixed baseline, fixed stereo setup) [2]. This discussion of resolution and geometry was helpful in choosing camera, baseline, and height parameters, which in our case was flexible depending the amount of area we wanted to observe and the depth resolution we required. Kevin Stefanik, et al., used a stereo setup to model terrain. They used a large remotely controlled helicopter as a platform. Their research used a baseline of 1.5m, and resulted in a depth resolution of 56-65cm at a distance of 40m [8]. This research, although used to map terrain, did not employ multiple UAVs and was not applied in a navigational sense.

3 Contributing Research

3.1 Theory

Of primary importance to the development of 3D reconstruction of terrain is capturing enough depth resolution. Another important factor to consider is the amount of area captured: in order to navigate, the 3D area captured needs to be wide enough to see an area unique enough to identify. All of these are linked by several more fundamental parameters, namely: baseline, camera resolution (both focal length and sensor resolution), altitude above the terrain, and overlapping image area. It is important to note that adjusting just one of these parameters has consequences on the rest of them; for example, if greater depth resolution is required, a longer focal length lens can be used. However, this also means that the overlapping image area is reduced and the overall captured area is reduced.

In order to begin narrowing these parameters down, a final capture area of about 1.5 x 1.5 km was decided upon. There was no empirical derivation of this figure, it is merely a figure that seemed both reasonable and large enough to capture unique identifiable terrain features by which to localize.

At this point it is advantageous to discuss the depth resolution equation for stereo cameras, seen below:

\[ \Delta z = \frac{z^2}{bf \epsilon_d} \]  

where \( \Delta z \) is the depth resolution in the center of the image (which is the location of maximum resolution, see K. Stefanik et al [10]), \( z \) is the altitude, \( b \) is the baseline, \( f \) is the focal length in pixels, and \( \epsilon_d \) is a disparity value. As can be seen, depth resolution is directly proportional to both baseline and focal length and quadratically dependent on \( \Delta z \). As in [2] by D. Gallup, et al., we have assumed \( \epsilon_d \) to be 1; this value also was reasonable in a calculation using baseline, focal length, and altitude values from K. Stefanik et al [8].

Although our cameras were simulated cameras, the parameters of the simulated camera could take on any value and therefore we modeled our simulated camera off a common 1280x900 3.75 \( \mu \) pixel resolution, 1/3” sensor format. Cameras of this resolution and format are common especially in computer vision and are inexpensive.

Using this sensor, a "normal" focal length (i.e., a field of view that replicates the perspective seen by the human eye) of 6.2mm (42.5630° x 32.6330°, approx. 1653 pixels), altitude of 2300m, and baseline of 300m, the resulting depth resolution \( \Delta z = 10.7m \) from (1). To examine the impact of sensor resolution and size on depth resolution, consult the table below ?? which keeps baseline, angle of view, and altitude approximately the same (and consequently overlapping image area).
<table>
<thead>
<tr>
<th>Sensor Dimension</th>
<th>Pixel Resolution</th>
<th>Pixel Size</th>
<th>Lens FL</th>
<th>$\Delta z$</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.2mm x 8.2mm</td>
<td>5496 x 3672 ($\approx 20$ Mp)</td>
<td>2.4µ</td>
<td>14mm</td>
<td>2.9m</td>
<td>Sensor based off of Sony IMX183 (1” type)</td>
</tr>
<tr>
<td>17.6mm x 13.3mm</td>
<td>4644 x 3506 ($\approx 16$Mp)</td>
<td>3.8µ</td>
<td>20mm</td>
<td>3.23m</td>
<td>Sensor based off of Panasonic MN34230 (4/3 type)</td>
</tr>
<tr>
<td>4.8mm x 3.6mm</td>
<td>1280 x 960 ($\approx 1.2$Mp)</td>
<td>3.75µ</td>
<td>6.2mm</td>
<td>10.7m</td>
<td>Sensor based off of Aptina MT9M034 (1/3” type)</td>
</tr>
</tbody>
</table>

Table 1: Impact of sensor size and resolution keeping baseline, focal length, and altitude constant

As expected, smaller pixels and larger sensor size can both increase resolution dramatically. However, these sensors are more expensive and the increased pixel count increases computation time.

To reiterate, the purpose of this experiment is to investigate GPS-less navigation for UAVs using 3D terrain data. The UAVs would need a height map of the area stored locally on board. The terrain beneath the UAVs would then be captured using nadir oriented cameras, one on each UAV. The distance between the UAVs would be the baseline of a large stereo capture. For the generation of disparity maps, the Semi Global Block Matching (SGBM) algorithm available in open source OpenCV packages was used, and OpenCV template matching using Normalized Cross-Correlation Matching Method (NCCMM) to find this area on a larger map of the entire area, from hereon out known as the global map. See Figure 1 below for a diagram representation of this process.

![Flow Chart of Localizing process](image)

Figure 1: Flow Chart of Localizing process

In applicable stereo imaging, there is a significant amount of image processing that the stereo pair must undergo before the disparity map can be calculated. For the highest quality depth maps images should be rectified before SGBM takes places and rectification requires processes of its own. First, the cameras should be approximately frontally parallel and horizontally aligned, which will make other computations less costly. After the cameras are configured in such a way, they need to be calibrated as a stereo pair. Once
the rectified images are rendered, the algorithm can create a depth map from the now row aligned images where each pixel of disparity contains information on how it relates to 3D space, namely its distance from the baseline. At this point, the local depth map created by the program is compared with the onboard global map using OpenCV’s template matching functions. Scale and orientation must be predefined by the user as template matching in OpenCV is not scale or rotation invariant. In this experiment, rotation and scaling were resolved by passing these parameters manually into the functions used to find the UAS locations. Once the image is found on the global map, the UAVs have been localized, as the relation between the center of the individually captured images represent the location of the UAV.

3.2 Simulated Terrain Experiment

This experiment used terrain generated in the computer program Terragen. A square area of 10 x 10 km was generated, and a ground truth grayscale heightmap was exported to serve as the global map (this can be seen in the appendix, Figure 4). Several different scenes and parameters were used. In general, it was found that a baseline of 300m at altitude of 2000m failed to produce a disparity map using SGBM with every scene tested. However, using the same scene and keeping all parameters constant, changing only the baseline down to 100m, a disparity map was successfully generated. This indicates that with a 300m baseline and 2000m altitude, there was not enough image overlap to successfully find disparities (however, this could be mitigated by a wider focal length lens or higher altitude, at the expense of depth resolution). The 300m baseline, 2000m altitude results in about 78% image overlap, while the 100m baseline, 2000m altitude results in about 92% overlap (the disparity mapping process may also reduce the final size of the disparity map). See Figures 2 and 8 below for the output of the failed disparity maps.

![Figure 2: Failed disparity map](image1)

![Figure 3: A second failed map](image2)

In total, we tested 8 renders of 5 scenes, which incorporated scenes with attributes ranging from distinct features such as hills to subtle features such as gentle valleys. It should be noted that the disparity maps that come out of the disparity generation program are very "flat", i.e., have no contrast and therefore need the pixel intensity values stretched. Ideally this should be done such that a pixel intensity value representing an elevation on the captured disparity becomes equal to the pixel value of an equivalent elevation on the global map. This stretching of pixel values will be discussed in application later in this paper. Additionally, a Gaussian blur was applied to these captured disparities to smooth the rough steps in elevation change. These processes were done visually in a photo editor. If there was any rotation with respect to the global map, this parameter was entered into the program manually. Finally the disparity map generated by the stereo pairs were scaled down before template matching processed the images. It is necessary that the spatial resolution of the captured disparity be similar to the spatial resolution of the global map. In this experiment, the captured disparities were downsampled to ten percent of their original sizes. This scaling does not need to be exact for the template matching to work; the renders we passed to the program captured different sized areas but all were scaled by 90%. The table below ?? shows an assessment of the different parameters on different scenes; Renders 1 and 4 below produced Figures 2 and 8 above. See Figure 5 in the appendix for the approximate locations of the scenes on our map.
Table 2: Results from 8 renders of 5 different scenes

In Table ?? above, every time a successful disparity was generated, the NCCMM template matching was able to correctly identify the location on the map (these disparity maps can be found in the appendix). This was case even in the more challenging scenes, with very few distinct features such as Scenes 2 and 5 (Renders 5 and 8). Additionally, Render 8 (Scene 5) had a reduced field of view compared to the other renders due to the longer focal length. This, in addition to the lack of distinct features, made this scene a worst case scenario in this experiment. However, a disparity was still generated and localization was successful. Render 7 was the widest area captured as well as the lowest resolution, with both a short focal length (4.18mm) and 2000m altitude. However, the correct mountain was identified despite similar size mountains being on the map. Even capturing a mountain by itself with no surrounding features, such as in Render 6, the correct mountain was identified. Comparing Renders 2 and 3, which give roughly similar captured areas, Render 3 gives more resolution as expected by the lower altitude; however, both were correctly localized. Again, it should be noted that although different sized areas were capture by the renders, they were all downscaled by 90% yet all successful renders were successfully located. The same is true of the rotation parameter discussed above; it need not be exact.

3.3 Discussion of Application

It is important to discuss that these simulations represent an ideal scenario. This experiment does not include potential problems with multi-UAV stereo imaging such as calibration of cameras, fixing the baseline between two flying craft (which might be accomplished with computer vision and edge detection), nonorthogonal baseline and direction of travel, et cetera. The cameras in this experiment exhibited ideal characteristics, such as no radial aberrations from the lenses, and might represent images after they have been rectified and aligned.

As was mentioned, if the scene captured had any rotation with respect to the global map, this deviation was entered manually into the program for this experiment. However, in application, the UAVs would be equipped with a magnetic compass and could just as easily pass a compass heading along with the stereo images into the software. The spatial scale ratio could be established pre-flight, when the cameras are calibrated.

Regarding the stretching of pixel intensity values this can be done systematically. For example, the UAVs would be flying at a constant altimeter setting, for example 2000m, and the terrain beneath would vary in elevation, changing distance to the UAV. When the stereo image is captured, a disparity value of the highest and lowest point can be output. Using stereo triangulation, which requires knowledge only of focal length, baseline, and disparity value, the distance to the terrain can be calculated (in effect, the elevation of the highest and lowest points can be determined). Assuming that the elevations and pixel intensity values on the
global map are known, this information can be used to assign the highest and lowest points in the captured image to the correct intensity values.

Finally, it must be mentioned that using grayscale height maps is not the only way to go about this comparison of captured terrain to known terrain maps; indeed, it may not even be the most desirable method. Most available terrain data is not in the form of grayscale height maps. Therefore, some work would be required to turn the available terrain data to a grayscale height map. Alternatively, the captured terrain data could be converted to whatever form is most convenient, for example contour elevation lines to compare to contour elevation maps.

### 3.4 Future Work

In addition to the real world application challenges mentioned above, there is much work to be done in this localization solution. More rigorous testing regarding what threshold of image overlap is required, what the minimum depth resolution is required, et cetera, is necessary. Additionally, more research needs to be done to investigate what the maximum size the global map can attain before it becomes too large to be useful in a computationally cost effective way, or before false positives become a problem. Similarly, more information is needed regarding the behavior of this method in the case of multiple very similar features in close proximity.

### 4 Conclusion

This paper presents a study of UAV homing in a GPS denied environment. More specifically it focuses on a system of drones, in this case two, which localize when GPS becomes degraded or non-existent. The results show that the two drones can each capture an image of the same ground terrain with significant overlap to be used as a stereo pair; in general, an initial overlap of 78% was too little, while an initial overlap of 92% was very usable. For analysis, 8 sets of 5 scenes were used to generate disparities. Using a Normalized Cross-Correlation Matching algorithm, the depth map was found on the global depth map and the location and orientation of the UAVs are calculated in every instance that a successful disparity map was generated. For future work there is the design of the method that will establish a baseline measurement between the two drones, as well as overcoming the challenges of multi vehicle stereo calibration. The extreme case thresholds need to be determined. Finally, future work could encompass path planning back home after 2010 localization occurs.
5 Appendix

Figure 4: Ground truth height map, which served as the map by which to navigate (10x10km)
Figure 5: A top view of the generated terrain, with approximate locations of the scenes marked (10x10km)
Figure 6: One of the images used for stereo imaging, Render 3. Left

Figure 7: One of the images used for stereo imaging, Render 3. Right
Figure 10: Disparity map for Render 5

Figure 11: Disparity map for Render 6
Figure 12: Disparity map for Render 7

Figure 13: Disparity map for Render 8
References Cited


