Photomosaic Mapmaking of the Seafloor using SIFT and Differential Lighting

Marcus Hammond
Aerospace Robotics Laboratory
Stanford University
marcus.hammond@stanford.edu

Abstract

This project will address the problem of creating mosaic maps of the seafloor using remotely-operated vehicles (ROV) or autonomous underwater vehicles (AUV) operated by the Monterey Bay Aquarium Research Institute (MBARI). The author believes that a number of shortcomings of the current system could be improved upon through the use of SIFT for image registration. Time permitting, the author would also like to investigate the use of differential lighting and shadow analysis to build up more detailed texture maps of the seafloor.

1. Introduction

MBARI currently performs mosaic mapping of the seafloor by running a remotely operated vehicle in a back-and-forth “lawnmower” path, grabbing frames from a video camera as it goes. The resulting images are correlated using Signum Laplacian of Gaussian (SLoG) filtering as described by Richmond [3]. This method is not robust to scale or orientation, and can only be done robustly with the high frame rate provided by video input. This is acceptable for performing registration in-line, but performs weakly when attempting to perform side-to-side correlation between two swaths. The current system requires offline verification of swath overlap, which is slow, and not robust to scale, orientation and illumination changes.

In this paper, the use of SIFT features ([2]) to perform the side-to-side swath correlation is explored. By using a small sample of robust image features to perform side-to-side correlation, the author hopes to speed up the process dramatically while maintaining or increasing the level of accuracy in the correlation.

2. Proposed Improvement

2.1. SIFT-based correlation and registration

SIFT features are robust to large changes in scale and orientation. Also, because each feature has a unique identifier, correlation can be performed between photos taken further apart in time and space. These are two instances in which SLoG correlation performs poorly. Once feature correlations are established, robust algorithms that reject outliers such as RANSAC or Hough transform can be used to register the images in the map. This can be seen in figures 1 and 2. Preliminary results also show drastic improvement in the number of side-to-side correspondences detected between swaths (figure 3).

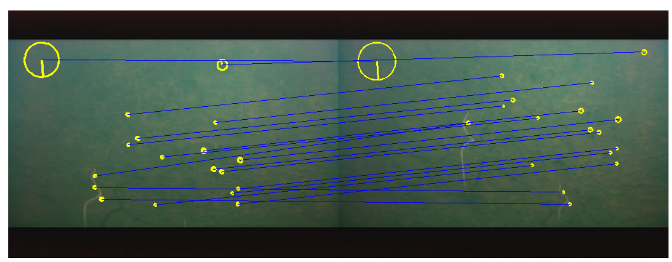

Figure 1. Though several false matches occurred between images, the majority of the SIFT matches show consistent frame-to-frame correlation

2.2. Building texture models with differential lighting

Although they are robust to scale and orientation changes, SIFT features are only partially robust to illumination changes. Because they use image gradient information, they are somewhat insensitive to global illumination changes, but much less so when faced with local illumination changes like shadows and changing light sources. Previous attempts to address this have included controlling lighting in a scene and the use of high dynamic range images. Rather than eliminate changes in illumination, the author would like to explore using multiple images taken

Figure 2. The RANSAC algorithm rejected the false positive matches, yielding a good transformation model between the two images in figure 1.

under different, controlled lighting conditions and extracting depth information about the seafloor features from the observed shadows. In-depth analysis of this technique will likely be beyond the scope of this course project, but the author hopes to begin exploring the idea.

3. Data Access

The ARL has mosaic data from a number of past dive missions that will be used for this project. The data includes imagery and odometry, as recorded by the MBARI submersibles' onboard IMU and Doppler Velocity Logger. The drift rate on these sensors is generally quite low, on the order of 1% of distance traveled, but this is still large enough to require vision based environment-relative position measurements.

For the differential lighting task, no data exists currently. If the project progresses to the point where this topic may be addressed, the author plans to manually take images using controlled lighting, trying to mimic seafloor conditions as well as possible. Looking to the future, it is likely that MBARI missions will be able to take this type of data.

4. Preliminary Results

4.1. Image correlation density improvements

As shown in figure 3, SLoG only produces a few side-to-side correspondences (red links). Using SIFT matching, correlations and mappings can be found between many more frames. This greater degree of interconnectedness between frames will be beneficial when performing bundle adjustment after all the data has been collected.

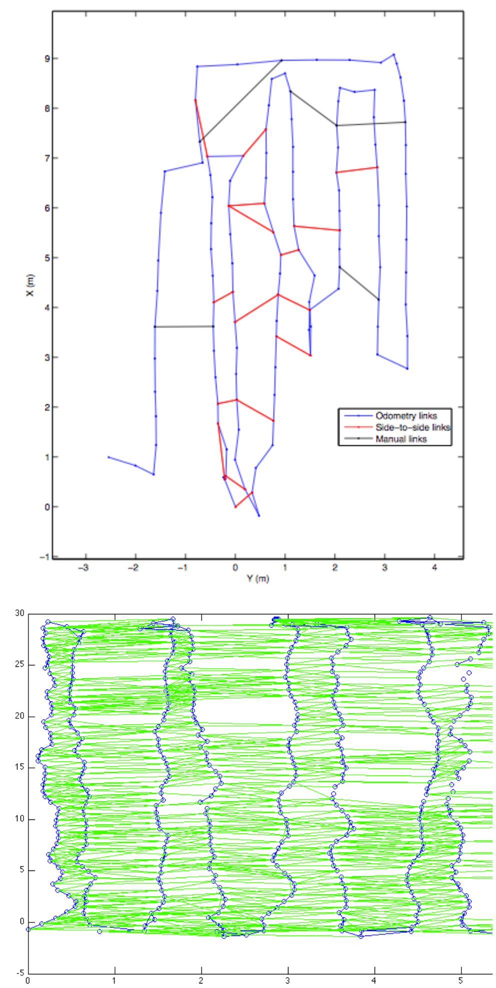

Figure 3. Side-to-side links using SLoG (top, Richmond [3]) vs. SIFT. SIFT has a much higher density of detections between frames.

4.2. Model selection for parameter estimation

Choosing a camera model is an important step in building a mosaic. Thus far, five different image transformation models have been tried: translation only, translation with in-plane rotation, translation with scaling, similarity transform, and affine.

While performing batch processing to minimize global reprojection error is an option for offline map optimization, it is also desirable to produce a map of reasonable accuracy in real-time as an ROV pilot aid while acquiring data. This consideration pushed the use of the above models. Other methods for image stitching similar to the one proposed use full perspective camera models, most notably the autostitch algorithm used by Brown and Lowe [1] However, his method assumes that the camera is purely rotated, making it possible to relate the transformations between successive frames through simple matrix multiplication. The large translations involved in mosaic mapmaking violate this assumption, and necessitate another approach. Using the transformations mentioned above, the global frame transformations can be calculated simply by multiplying their between-frame transformation matrices.

Figure 4. The most successful model so far has been the simplest, where only two parameters, x and y translation are estimated

The preliminary results are somewhat surprising. The most successful method is the one in which only the x and y translation parameters are estimated. This can be seen in figure 4. The author believes there is unmodeled pitch and possibly yaw bias in the camera that causes apparent looming and strafing motion in the field of view that is not uniform throughout the image. Adding degrees of freedom to the model allow it to try to account for this, but since the error is not zero-mean, it compounds.

Possible solutions for this are to try to estimate these parameters and predistort each image before extracting SIFT features and performing correlation, in effect "knocking out" the error so it cannot propagate.

4.3. Visualization tool for real-time pilot aid

One goal of this project is to give ROV pilots a tool for gauging how well they are achieving coverage of a given section of sea floor in real-time. Figure 7 shows one conception of such a tool. This tool will allow MBARI to create maps with more confidence and autonomy from the ARL.

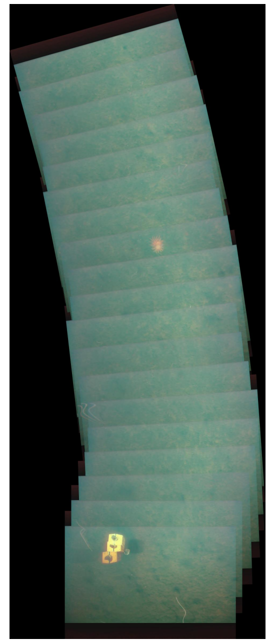

Figure 5. When in-plane rotation ψ is also estimated, the swath begins to drift. The author suspects that a pan-tilt camera bias is causing this behavior.

Figure 6. The affine model also falls victim to the distortion seen in the rotation model, though its manifestation is different. Each successive frame suffers from a greater degree of skew, rotation, and scaling.

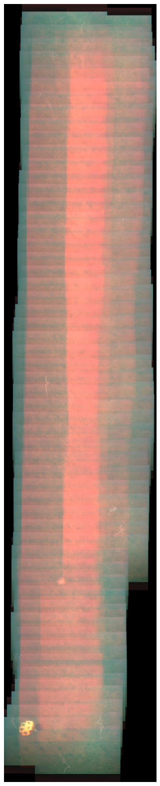

Figure 7. This map was generated in real time using a 2-parameter model. Areas of overlap were highlighted by boosting the red values of the pixels by a small amount. The effect is additive, so areas of multiple frame overlap are more red. This is intended for use by ROV pilots to show them in real time how well they are achieving coverage.

References

[1] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision*, 74(1):59–73, 2007.

[2] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60:91–110, November 2004.

[3] K. Richmond and S. Rock. An operation real-time large-scale visual mosaicking and navigational system. *OCEANS*, pages 1–6, September 2006.