

8-2001

Improvement of Image Alignment Using Camera Attitude Information

Yuri Rzhanov

University of New Hampshire, Durham, Yuri.Rzhanov@unh.edu

Lloyd C. Huff

University of New Hampshire, Durham

Randy G. Cutter Jr.

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/ccom>

 Part of the [Oceanography and Atmospheric Sciences and Meteorology Commons](#)

Recommended Citation

Rzhanov, Yuri; Huff, Lloyd C.; and Cutter, Randy G. Jr., "Improvement of Image Alignment Using Camera Attitude Information" (2001). *International Symposium on Signal Processing and Its Applications (SPIA)*. 227.
<https://scholars.unh.edu/ccom/227>

This Conference Proceeding is brought to you for free and open access by the Center for Coastal and Ocean Mapping at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Center for Coastal and Ocean Mapping by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

IMPROVEMENT OF IMAGE ALIGNMENT USING CAMERA ATTITUDE INFORMATION

Y. Rzhanov, L. Huff, G. R. Cutter

Center for Coastal and Ocean Mapping (C-COM)
University of New Hampshire
Durham 03824, USA

email: yuri.rzhanov@unh.edu, lhuff@cisunix.unh.edu, gcutter@cisunix.unh.edu

ABSTRACT

We discuss a proposed technique for incorporation of information from a variety of sensors in a video imagery processing pipeline. The auxiliary information allows to simplify computations, effectively reducing number of independent parameters in the transformation model. The mosaics produced by this technique are adequate for many applications, in particular habitat mapping. The algorithm is demonstrated through simulations and hardware configuration is described in details.

1. INTRODUCTION

The ideal product from seafloor imaging efforts would be a complete 3D reconstruction of the scene with the resolution required to resolve essential details, and accurate georeferencing such that subsequent maps could be compared directly. Essentially, that would provide real seafloor images from any perspective at the finest scales (the basin with the water removed). To accomplish that, accurate bathymetric data must be acquired first, the terrain reconstructed, then optical imagery draped over. The scene would then be ready for virtual reality applications and 3D visualization which could provide additional insights with innovative analyses. This technique would definitely find many applications in marine biology, geology, and underwater archaeology. Implementation is the difficulty.

1.1. Constraints and alternatives

While some steps already have been taken in that direction [see, for example, [1]], field data collection for the approach requires highly stable platform (submarine, AUV), expensive microbathymetric equipment and precise underwater navigation.

However, if it is chosen not to use bathymetric methods (due to lack of time or resource constraints), the optimal product for underwater imaging is a site map - projection of the site features onto two dimensional planar surface. In this case we do not try to deal with effects of bathymetry

- instead we are compensating for it, as if the scene was indeed flat.

In the present paper we concentrate on a technique allowing construction of such a map. Our goal is to construct a view with as few distortions as possible, so the relative positions of scene features (objects) can be accurately estimated and measured.

We show that this can be achieved by inexpensive means using consumer-available products, such as a hand-held video camera and a set of attitude sensors, and that the product is valuable for seafloor mapping.

2. SPECIFIC APPLICATIONS

One primary reason for collecting bottom video imagery is to characterize seafloor habitats or provide information for ground-truthing acoustic data. Video imagery provides data relating substrate characteristics, organisms (density, spatial patterns and variability), and associations of organisms with substrate. This data can reliably characterize many benthic, epibenthic and, less accurately, fish populations. Although video imagery does not provide direct information about habitat utilization, it can sometimes be inferred since many benthic and fish species tend to associate with and utilize particular bottom types and substrate characteristics for shelter, nutrition, or reproduction during some stage in their lives (e.g. [2]). Because of that, mapping the substrate and biological features over a range of scales is important for habitat surveys, and video mosaics provide the most complete and accurate representation of the substrate from scales of millimeters to hundreds of meters.

2.1. Difficulties with underwater video imaging

Habitat mapping is most important in estuaries, coastal, and continental shelf regions because of the reliance of human-exploited fisheries species upon these areas. Unfortunately conditions in these parts of the oceans are worst for optical imaging because of suspended sediments, particles and planktonic organisms, all of which impede optical paths and

scatter or absorb light. In addition, underwater scenes are much less accessible than those on the land - investigators have to wear diving equipment or be restrained by an underwater vehicle, or remotely operated tools must be used.

Even when good quality video is acquired, analysis is limited to either review of the video recording or examination of individual still images constructed from video frames. In either case, sample time interval and travel distance must be maximized to avoid duplicating counts, and minimized to avoid missing areas. These problems can be overcome using sequential video frames to construct a video mosaic map from which multi-scale measurements can be made (Fig. 1).

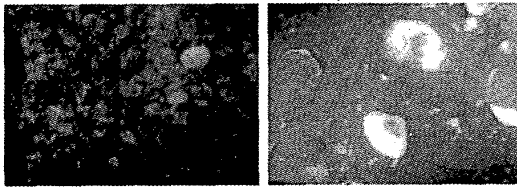


Figure 1: Examples of frames from underwater video.

2.2. Virtues of a video mosaic record

Mosaics provide continuous data. Both broad areal coverage, by multiple track overlap, and narrow transect coverage provide direct information at scales larger than point sample data. Point sample data can be used only to estimate intermediate spatial scales, larger than the field of view in an image for our case. Mosaics allow multi-scale analysis without the possibility of duplicate counts. Counts from mosaics may suffer from distortions caused during mosaicing, however population estimates will be more accurate than those made from counts of randomly selected frames. Although randomly selected frames may be more reliable for identifying certain small features, population estimates will incur error unless the population follows an idealized distribution that is known *a priori*. Many benthic and epibenthic populations have aggregated distributions, such that they are dense in a few areas but sparse for most, and therefore are unlikely to be enumerated accurately using single image selections. That is especially true for the case of larger organisms. In addition to representing the organisms and features, mosaics also allow distances between features to be measured directly.

2.3. Video image mosaicing

In the scope of this paper we assume that the scene we want to map is flat. Errors in the mosaic due to terrain effects errors in the mosaic may be neglected for our application as

there are no large errors (seafloor areas mapped using devices employing optical imagery are assumed to be flat) and small local distortions cannot affect the biological feature data. The analysis errors incurred because of flat-bottom assumption are less than those related to navigational inaccuracies that must be considered if separate frames were considered.

In order to generate the product inexpensively, we employ either a diver-held or towed video camera in an underwater housing. Both camera platforms have 6 degrees of freedom in their motion. Translation, yaw, and heave do not pose problems for map construction. These motions are described by an affine model and all four parameters of the model (scaling, rotation and translation in two orthogonal directions) can be robustly estimated by automatic featureless Fourier transform-based technique [3, 4]. However, any tilt of the camera poses a serious complication. Perspective distortions of the acquired images (due to pitch and/or roll) combined with translation of the point of view, require more complex transformations such as 8 parameter projective transformation [5, 6]. There are two means of recovering these parameters: (1) manually choosing four (at least) tie points in each image to be co-registered (perhaps by pre-arranged insertion of features into the scene) and solving eight equations for eight unknown parameters; or (2) formulation of a non-quadratic minimization problem for pixel value differences and consequent numerical solution using Levenberg-Marquardt [1, 6] or Gauss-Newton scheme [7]. Both methods have drawbacks: human intervention is required in the first, and in second case intensity changes (motion of the light source) and objects movements make optimization methods error prone. Besides, optimization techniques require good initial guess, and if the latter is chosen inappropriately convergence of the algorithm is not guaranteed.

In this paper we propose to use additional information from consumer-grade attitude sensors to effectively reduce the complexity of the transformation model: from projective to affine. Estimate of the tilt angle allows to reproject perspective distorted images onto the used defined plane. Reprojected images are related to each other by an affine model. The latter can be efficiently solved with robust automatic non-iterative methods, has guaranteed convergence and is not influenced by inhomogeneous lighting. By means of simulating video acquisition with significant tilt angles, we show that although acquired frames do not appear visually distorted, they cannot be reliably assembled in a mosaic. However, if tilt angles can be estimated, even inaccurately, frames can be corrected for projection distortion and the quality of the mosaic is improved dramatically. Simulations allow the establishment of a relationship between errors in tilt angle measurement and quality of the produced mosaic.

3. INSTRUMENTATION

The video camera used for this work is the Sony DCR-TRV310 Digital Video Camera. It is mounted in a diver held underwater housing manufactured by Ikalite. The minimum auxiliary data to support the improved mosaicing process is a time series of camera roll and pitch that is synchronized with the video frames. The synchronization must be maintained at the frame level, even though successive frames in a video are virtually never employed in developing a mosaic (because not all of them are necessary). Simulations of the proposed roll/pitch corrected mosaicing process indicated that the accuracy of the roll/pitch information should be on the order of one degree, or better. The benefit of using roll/pitch information that is better than one degree does not justify the instrumentation cost to achieve an accuracy of 0.1 degree, rather than one degree. This is especially true in light of other factors like pixel resolution and the losses in fidelity of the images, associated with rotating the video frames into a common, north-up frame of reference.

The Precision Navigation Model TCM2-50 was selected for the required roll/pitch information. Its size and power requirements are well suited for inclusion within the camera's underwater housing. The roll/pitch accuracy of the unit in a static environment is 0.4 degree. An important consideration in the choice of the TCM2-50 is that it also provides an azimuth relative to magnetic north, which is accurate to 0.5 degree in a static environment. The processing algorithm is capable of determining rotations between successive frames, although any sequence of rotations is subject to cumulative, random walk errors. Camera orientation relative to magnetic north can be utilized to develop the mosaic in a north-up reference and to constrain the algorithm's determinations of frame-to-frame rotation. The TCM2-50 also has the desirable characteristic of being programmable with respect to sample rates, averaging times, and output intervals. It is intended to output data twice per second, that are one-second averages of the parameters, which have been sampled at 16 Hz.

The RS232 formatting of the TCM2-50 output is an essential element for recording the heading, roll, and pitch information in synchronization with the video frames. The TCM2-50 output is reduced in amplitude and ac coupled into the audio input channel of the video recorder. Internal to the Sony DCR-TRV310, the signal input is digitized to 12 bits at a bit sample rate of 32 kHz and recorded onto the tape, in synchronization with the video imagery. The heading, roll, and pitch data are automatically retrieved through reconstruction, detection, and decoding the time series waveform of the RS232 output message.

This videotape technique of recording/retrieving RS232 messages in synchronization with video imagery has been successfully demonstrated. Figure 2 presents part of the

reconstructed time series of a navigational position output message from a GPS receiver that had been recorded on videotape. Videotaped RS232 messages are automatically recovered with less than one bit error per million.

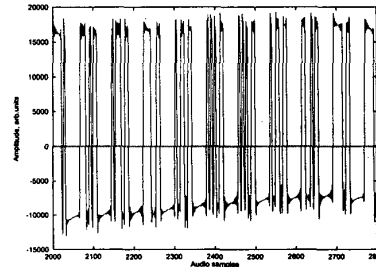


Figure 2: Time series of the GPS message.

4. RESULTS

Figure 3,a shows square frames generated from a raster image with inclined pinhole camera model (pitch equals 10 degrees). Distortions are not visually noticeable, but the mosaic constructed from these frames has significant curl (Fig. 3,b). (Reasons for this curl and ways to correct the mosaic are discussed in [8]). Correction of frames for pitch (Fig. 3,c) allows for creation of an almost ideal mosaic with no curl at all (Fig. 3,d).

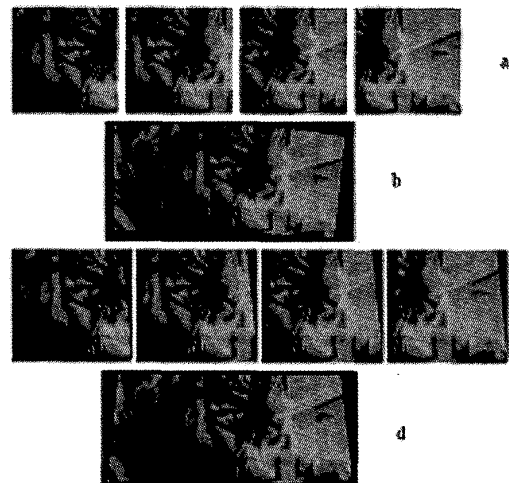


Figure 3: Separate frames used in simulation (a, c) and resulting mosaics (b, d).

Application of the technique to the mosaicing of video

imagery collected underwater (even with inaccurate determination of the pitch angle) produces mosaics of high quality, as the one shown in Fig. 4.

5. CONCLUSIONS

This paper has presented a technique for video mosaicing of underwater images. Auxiliary information from attitude sensors has been incorporated into the algorithm to simplify image processing and to assist in creation of mosaics. The method has been applied to marine biological measurements and provides a cost-effective alternative to previously reported methods. The technique shows much promise for gathering new types of information from the seabed using video as opposed to traditional acoustic imaging.

6. REFERENCES

- [1] H. Singh, L. Whitcomb, D. Yoerger, O. Pizarro, "Microbathymetric mapping from underwater vehicles in the deep ocean." *Computer Vision and Image Understanding*, vol. 79, pp. 143-161, 2000.
- [2] D. B.Packer, T. Hoff, "Life history, habitat parameters, and essential habitat of mid-Atlantic summer flounder." *AFS Symposium 22*, pp. 134-147, 1999.
- [3] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation and scale-invariant image registration." *IEEE Transactions on Image Processing*, vol. 5, pp. 1266-1271, 1996.
- [4] Y. Rzhano, L. M. Linnett, R. Forbes, "Underwater mosaicing for seabed mapping." *2000 International conference on Image Processing, ICIP2000, Vancouver, Canada*, vol. 1, pp. 224-227, 2000.
- [5] S. Mann and R. W. Picard, "Video orbits of the projective group: a simple approach to featureless estimation of parameters." *IEEE Transactions of Image Processing*, vol. 6, pp. 1281-1295, 1997.
- [6] H. S. Sawhney and R.Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 235-243, 1999.
- [7] R. Radke, P. Ramadge, "Efficiently estimating projective transformations." *2000 International conference on Image Processing, ICIP2000, Vancouver, Canada*, vol. 1, pp. 232-235, 2000.
- [8] A. Zomet, S. Peleg, C. Arora, "Rectified mosaicing: Mosaics without the curl." *IEEE Conference on Computer Vision and Pattern Recognition, CVPR2000, Los Alamitos, USA*, vol. 2, pp. 459-465, 2000.

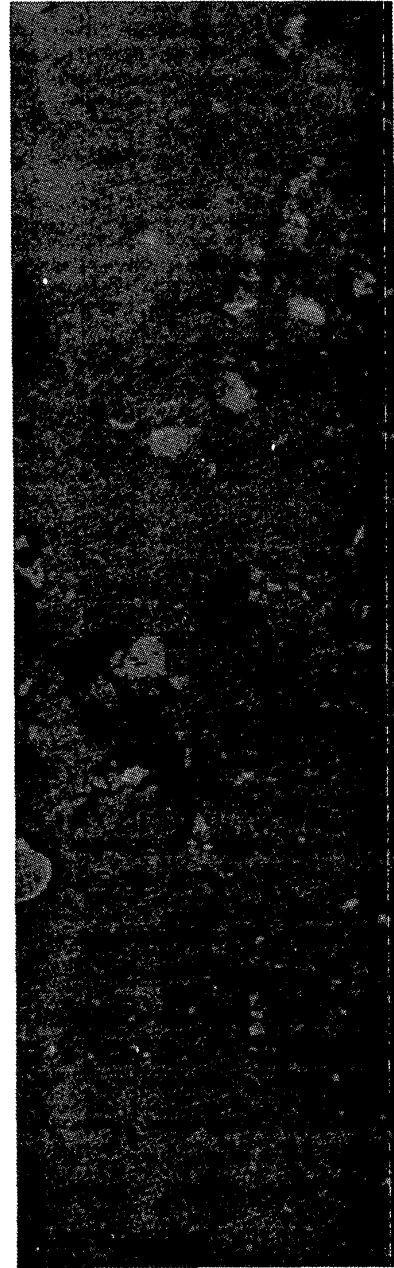


Figure 4: Mosaic of approximately 100 frames.