11-2004

Deep-sea image processing

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

Larry A. Mayer
University of New Hampshire, larry.mayer@unh.edu

D J. Fornari
Woods Hole Oceanographic Inst.

Recommended Citation
Rzhanov, Y.; Mayer, L.; Fornari, D., "Deep-sea image processing," in OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04 , vol.2, pp.647-652 Vol.2, 9-12 Nov. 2004. doi: 10.1109/OCEANS.2004.1405498

Deep-sea Image Processing

Y. Rzhanov, L. Mayer, Center for Coastal and Ocean Mapping

University of New Hampshire

Durham, NH, 03824, USA

yuri.rzhanov@unh.edu

larry.mayer@unh.edu

D. Fornari, Geology & Geophysics Dept.

Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

dfornari@whoi.edu

Abstract **- High-resolution seafloor mapping often requires optical methods of sensing, to confirm interpretations made from sonar data. Optical digital imagery of seafloor sites can now provide very high resolution and also provides additional cues, such as color information for sediments, biota and divers rock types.**

During the cruise AT11-7 of the Woods Hole Oceanographic Institution (WHOI) vessel R/V Atlantis (February 2004, East Pacific Rise) visual imagery was acquired from three sources: (1) a digital still down-looking camera mounted on the submersible Alvin, (2) observer-operated 1- and 3-chip video cameras with tilt and pan capabilities mounted on the front of Alvin, and (3) a digital still camera on the WHOI TowCam [1]. Imagery from the first source collected on a previous cruise (AT7-13) to the Galapagos Rift at 86°W was successfully processed and mosaicked post-cruise, resulting in a single image covering area of about 2000 sq.m, with the resolution of 3 mm per pixel [2]. This paper addresses the issues of the optimal acquisition of visual imagery in deep-sea conditions, and requirements for on-board processing. Shipboard processing of digital imagery allows for reviewing collected imagery immediately after the dive, evaluating its importance and optimizing acquisition parameters, and augmenting acquisition of data over specific sites on subsequent dives.

Images from the DeepSea Power and Light (DSPL) digital camera offer the best resolution (3.3 Mega pixels) and are taken at an interval of 10 seconds (determined by the strobe's recharge rate). This makes images suitable for mosaicking only when Alvin moves slowly (<<1/4 kt), which is not always possible for time-critical missions.

Video cameras provided a source of imagery more suitable for mosaicking, despite its inferiority in resolution. We discuss required pre-processing and image enhancement techniques and their influence on the interpretation of mosaic content.

An algorithm for determination of camera tilt parameters from acquired imagery is proposed and robustness conditions are discussed.

I. INTRODUCTION

In recent years optical methods have gained popularity in deep-seafloor studies. On one hand, they guarantee resolution that is usually far superior to any acoustical method; on another, they allow one to quickly explore relatively large areas that cannot be possibly covered by an "ultimate groundtruthing method" - core sampling. The human visual system, which is a massively parallel processor, provides the highest-bandwidth channel into our cognitive centers through direct observation of the seafloor morphology and organisms inhabiting the seafloor. These types of observations provide tremendous insight in understanding biological, geological and chemical processes operating in the deep sea.

This realization has stimulated researchers to develop digital technologies and lighting systems suitable for operation in the deep ocean. Unfortunately, severe attenuation of visible light and limited power capabilities of many submersible vehicles, require acquisition of imagery from short ranges, rarely exceeding 8-10 meters. This means that although modern video- and photo-equipment makes high-resolution video-survey possible, the field of view of each image remains relatively narrow.

To compensate for this deficiency, researchers have been developing techniques allowing for combining images in a bigger picture - mosaicking (e.g. [3]). A properly constructed accurate mosaic has a number of well known advantages in comparison with the original sequence of images, the most notable being improved situational awareness. A trained observer may be able to keep in memory a few prominent features during an hour-long survey, but if a vehicle track has just several turns, the observer would not be able to judge how one feature is oriented with respect to another. Another advantage is that a mosaic dramatically simplifies the search for some particular feature. Storage of a mosaic image is significantly less demanding than that of the image sequence, as the redundancy has been removed. As an additional bonus, mosaicking process allows for detection and extraction of objects moving with respect to the background.

2. EXPERIMENTAL DATA

In this paper we address mosaicking issues with respect to the deep-sea imagery collected by the deep submergence vehicle Alvin, operated by WHOI. On a typical dive, Alvin has three 1- and 3-chip video cameras mounted on the front of the vehicle, with pan and tilt capabilities, operated by a pilot and observers. The illumination is provided by 2-4, normally 400 watt HMI (metal-halide) lights pointed forward and down, mounted on the front of the vehicle.

Alvin also has a digital down-looking camera with a strobe illumination (600 watt/sec, total from two heads),

0-7803-8669-8/04/$20.00

sufficient to adequately illuminate the seafloor from an altitude of 6-7 meters. Images taken from higher altitudes do not have enough quality for identification of essential features. After the dive is finished, still images are combined in a gallery; footage from all video cameras is stored on DVCam tapes.

The repetition rate of images from the DSPL digital camera (10 seconds) is limited by the strobe recharging process. Combined with the limitation on an altitude, for images with 45 degrees field of view (along the motion direction) to have overlap of 65 percent (which is required for automatic registration technique) the vehicle speed must be less than 0.175 m/s. For lower altitudes this limitation becomes proportionally stricter. Most of the missions are time-critical, so it is apparent that quite often the collected imagery cannot be processed automatically. This was the case with the digital image data collected on the AT7-13 cruise to the Rosebud hydrothermal vent field [2]. This site was discovered in May 2002 in the Galapagos Rift near 86° 13.5'W during a series of Alvin dives and ABE autonomous vehicle surveys [4]. A complete survey of the Rosebud vent site was carried out on Alvin dive 3790. Submersible position was determined by integration of 1.2 MHz bottom-lock Doppler sonar velocity data logged at 5 Hz, integrated with heading and altitude data from a north-seeking fiber-optic gyroscope at 10 Hz, and initialized with a surveyed-in long baseline transponder navigation system providing geodetic position fixes at 15 s intervals. About 700 images were found to constitute a non-interrupted sequence that was possible to assemble in a single mosaic. However only less than 30 percent of all image pairs had overlap sufficient for automatic co-registration. The main bulk of the images required manual feature extraction and subsequent calculation of transformations relating adjacent images. (Transformation is a law determining how the pixels of one image map onto the pixel space of another image.)

2. DATA PROCESSING

The onboard mosaicking process consisted of three main stages: pre-processing, pairwise image co-registration and global alignment. The high quality lens system and optics on the DSPL camera did not introduce distortions worth correcting, so images only underwent histogram equalization which removed effects of inhomogeneous illumination and enhanced contrast.

Partial mosaics allowed us to determine substantial overlap between non-sequential frames, and the corresponding transformations were found manually via feature extraction. All found transformations, relating consecutive and non-consecutive images in the sequence, were then submitted to the global alignment procedure. The process of building a final mosaic is somewhat similar to numerical integration - errors incorporated in transformations relating images force the track to wander away from its true position (drift). Obviously the choice of the camera motion model (i.e. type of transformation) affects the way errors are accumulated.

While the simplest, 2-parameter translational model allows the track to deviate in two dimensions, an 8-parameter perspective model that in general describes much more complex motion including tilts and rotations, may cause the track to perform spectacular overturns. Even more sophisticated models attempt to simultaneously solve for camera motion and terrain bathymetry, and the error accumulation in these cases lead to the appearance of various unrealistic trends and slopes [5].

As a model of choice we decided to use the rigid affine model (RAM), that is complex enough to describe camera translation in all three dimensions and rotation about its optical axis (e.g., [6]). At the same time this model has a number of important advantages that will be pointed out in the explanations below.

It is convenient to present RAM as a particular case of the perspective model, with corresponding transformation described as 3x3 matrix. Cascading the transformations would then correspond to simple matrix multiplication. However as RAM has only four parameters, the corresponding matrix has two elements equal to zero, and two additional constraints on other elements. The ninth element, as usual, is equal to one, which reflects the scaling ambiguity of 3D-to-2D mapping.

One of the advantages of using RAM is a reliable automatic method of finding of transformation coefficients. We have previously described this method in detail [7]. Another advantage is the possibility to use an *a posteriori* scheme of transformation quality assessment. Estimation of the transformation quality presents a challenge for any model. The most reliable objective technique is a calculation of an average per-pixel error, that employs the so-called "brightness constancy constraint". This technique cannot be used when data are collected with artificial illumination, and shadows and highlights are changing dramatically from one image to another. Thus, goodness of found transformations can be verified only by human observation, which is both subjective and time consuming. For RAM, we have developed an artificial intelligence scheme based on the Support Vector Machine [8, 9] that provides a quantitative estimate of co-registration success. An application of this scheme alleviates some of burden from the human operator, allowing concentrating on marginal cases, and provides weight coefficients for global alignment stage (see below).

Non-sequential transformations ("cross-links") impose additional restrictions on the relationship between different images (or "world" transformations, describing how particular images are being mapped onto chosen common imaging surface) and help limit errors. In general, all coefficients of all transformations are used to construct a penalty function that has to be minimized to find an optimal solution. Without cross-links, the optimal solution can be calculated by simple cascading of the transformations, and the corresponding value of the penalty function is zero. With cross-links added, optimization can be performed by, for example, the Levenberg-Marquardt technique, that requires iterative solution of a sparse nonlinear matrix equation of order 4(N-1)x4(N-1), where N is the number of images. With a significant number of images this becomes a formidable task, but here, RAM offers another advantage - it allows one to decouple variables responsible for rotation and vertical translation of camera from those responsible for horizontal translation. Global alignment thus splits into three independent steps: the first and second involve inversion of (N-1)*(N-1) matrix, and the third a: 2(N-1)*2(N-1) matrix.

Any closed loop chosen among available transformations has a property that the total rotation of the camera along the loop is a constant equal to multiple of 2*Pi (loop topological constant). Even more, it can be shown that if the chosen loop is minimal (i.e. does not consist itself of two or more smaller loops), this constant can be only $+2*\pi$,

0, or -2*π. Even for inaccurate data it is straightforward to calculate the topological constant for any loop.

The same consideration is applicable to vertical translation of a camera (or more accurately, camera altitude), that is directly related to a scaling coefficient in the RAM. A scaling coefficient (equal to ratio of altitudes for respective images) is a multiplicative variable, so the product of scaling coefficients along the loop is equal to one. (As the imaged surface is considered to be flat, there are no topological constants in this case.) Both rules formulated above are equivalent to the well known Kirchhoff's law for voltages. To perform these stages of global alignment, the camera track is divided in minimal loops (each loop must obey the rules formulated above). The results of the first and second stages of global alignment are lists of relative transformations (similar to those calculated from co-registration procedure). The last stage has to be done for the "world", or absolute, transformations. The penalty function consists of contributions from all relative transformations (both consecutive and cross-links). In the beginning, consecutive links contribute nothing, because absolute positions have been calculated using only consecutive links. Large contributions come from cross-links, due to error accumulation. In the optimization process, some compromise between both groups of transformations is found, and relative transformations are recalculated again using optimal absolute transformations.

Convergence of all stages is slow. For example, global alignment of 700+ images for the mosaic presented in [2] required ~2 hours on a PC with an Intel4 3 GHz processor.

Different types of challenges are faced when the imagery is coming from a video camera. For video camera overlap between consecutive frames is close to 100 percent, so in order to reduce necessary processing time video sequences have to be decimated. Depending on the speed of the vehicle we reduced the data to 1-3 frames per second (from the standard 30). However this advantage is "compensated" by significantly lower resolution than the one for still images and, most importantly, by poor camera positioning. All three video cameras on Alvin have a primary objective to assist in direct observation by the scientists in the submarine. Hence the cameras are mounted relatively high above the seabottom and are tilted, so that the view is not obscured by Alvin's frame. The cameras and pan/tilts are operated by scientists and Alvin pilot, so the recording channel can be switched, and tilt is varying in time. This practically precludes possibility to combine video images in a multipass mosaic. Even highly recognizable relatively flat features look in a mosaic dramatically different when approached from another direction. Nevertheless single-swath mosaics were very useful during the AT11-7 cruise, especially for long runs with constant camera orientation.

As mentioned above, multipass mosaics could be corrected for curl and scaling by employing the properties of closed loops. Linear transects do not have this advantage. As a result, slight camera tilt (or seabottom slope) causes a mosaic built only on the basis of imagery to curl, often deviating dramatically from the true course (Fig. 1).

Similarly, camera pitch causes an artificial scaling effect. If the camera is pitched forward, any object in a field of view appears larger in the near range than in the far range. As a result, the mosaicked image will be decreasing in width, as if having perspective distortion (Fig.2). If the camera was known not to change altitude during a transect, it is sensible to ignore scaling factor altogether. However, in this case all height variations of the seabottom are ignored too, and the mosaic may have serious distortions and be not suitable for inclusion in a GIS database.

The possible alternative is to first build mosaic on the basis of imagery only and then to utilize available navigation and attitude information for the camera for mosaic modification (warping). The first step guarantees the optimal continuity of mosaic; while the second step provides for conformance with known auxiliary information. While a mosaic is being built, positions of pixels corresponding to centers of transverse sides are recorded in a separate file. Corresponding positions in the real units are calculated from camera position, heading and altitude at the moment of image acquisition. Mosaic resolution (scale between pixel and real spaces) is estimated from known camera altitude for the start frame. Both sets of points (obtained from imagery and estimated from navigation) are used to calculate coefficients for 2D thin plate spline [10, 11], that are later used for warping the mosaic to a required shape.

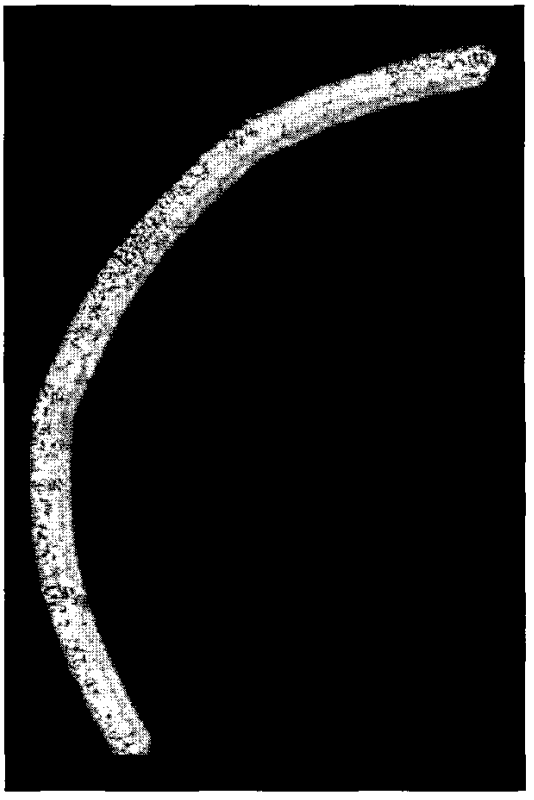

Fig. 1. 30 meter long transect demonstrating strong curl due to camera roll.

During the dive, all three video sources were recorded on digital DVCam tape. Observers had an option to show (and record) on-screen information (overlay) with Alvin's position and orientation. This display is useful for observation, however it reduces area on the frame that can be used for mosaicking. Immediately after each dive observers selected clips with the most scientifically interesting footage (based on their notes, or reviewing video footage). Typically, around 6-7 clips were chosen, each representing 5 to 8 minutes of dive time. These clips were subsampled to the required rate

of 1-3 fps, to reduce consecutive overlap to the optimal 65-80 percent. After sub-sampling, the frames were cropped from original 720x480 pixels to 688x376 pixels, to remove the data overlay from the top part of the screen and black padded margins (result of conversion from DVCam tapes to digital video files in AVI format). For fast processing, frames were then further reduced in size by a factor of 2, to 344x188 pixels using Lanczos filtering. Even when illumination was sufficient to visually extract main features, we found that it is was helpful to enhance the contrast by applying adaptive histogram equalization. We have used a contrast limited adaptive histogram equalization technique [12]. This algorithm is designed to work on grayscale images and distorts colors of the acquired footage. Nevertheless we have retained the color as it appears to help in geological and biological interpretation.

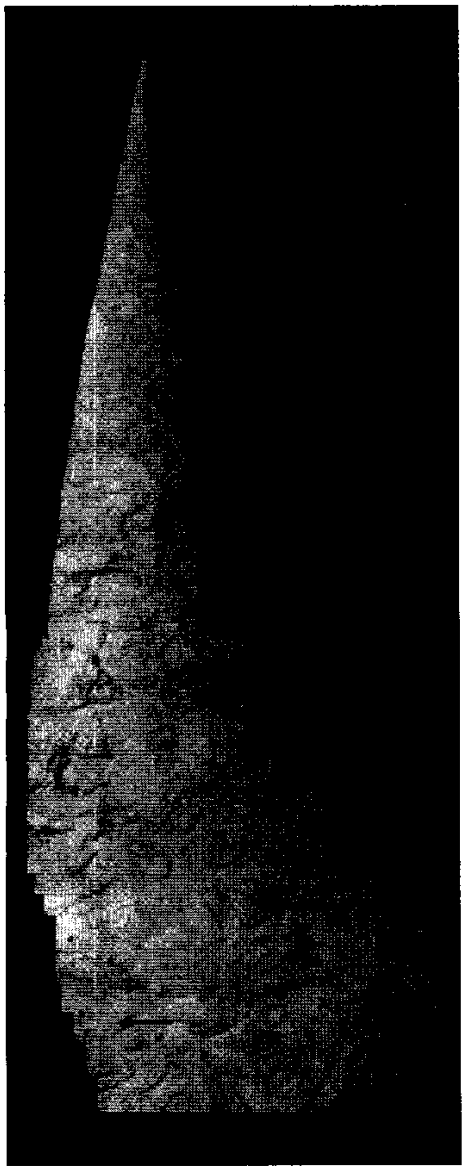

Fig. 2. Tilted (pitched) camera causes mosaic to exhibit perspective distortion.

Automatic registration of two frames took approximately 5 seconds on a PC with an Intel4 3 GHz processor. Typical sequences of 500 frames (~8 min of footage) required about 40 min to process. Quality assessment based on an AI scheme described above allowed the operator to quickly check continuity of the processed clip. Typically, bad registrations were detected when video channels were switched ("cut"), imaged terrain had exceptionally strong 3D content, or a moving object (for example, Alvin's manipulator) appeared in the camera's field of view. Human intervention was needed to resolve split sequences (first case), to approve the transformation marked as bad (second case), or to cut off footage that cannot be mosaicked (third case).

3. PROBLEMS TO BE ADDRESSED

Positioning information collected during the dive was usually reprocessed afterwards, by merging LBL fixes (once in 15 sec) with the DVL data (Fig. 3). However for some dives LBL data was not available. Hence the positioning information was sufficient for correctional warping of mosaics for short transects, but not for inclusion of mosaics in GIS database.

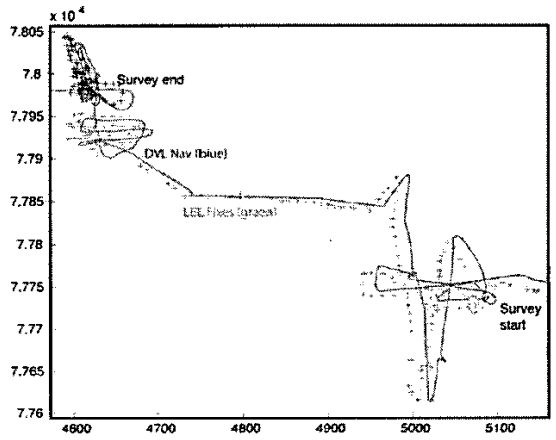

Fig.3. Example plot showing Alvin LBL (green) and DVL (blue) navigation data for dive 3976.

Currently we assume that the positioning sensor has the same location as the video camera. This assumption is not accurate, and offsets are in the order of meters in all three dimensions. Moreover, camera tilt shifts the location of center of the frame even further, and this shift depends on vehicle altitude as well as actions of observer controlling this camera. These actions are not being recorded at present, so the uncertainty in mosaic geo-referencing due to camera attitude may be up to 5-10 meters (i.e. on the same order as positioning uncertainty for LBL navigation).

In the future we plan to use only footage from a dedicated video camera, with fixed (vertical) orientation, and known offsets with respect to vehicle sensors.

4. TILT DETERMINATION FROM IMAGERY

Often the seafloor can be considered relatively flat (in comparison with the camera altitude), and quality of constructed mosaics can be improved by rectifying acquired frames prior to registration. However the pan/tilt cameras' attitude is not currently measured on Alvin. Hence tilting

angle has to be guessed or estimated from the imagery. We have developed a method to estimate tilt parameters for transects satisfying the following conditions: (a) tilt is constant, (b) vehicle is moving in a straight line, and (c) seabottom is relatively flat.

Transformation relating image acquired by a normally-looking camera, with image from the camera with roll R, pitch P, and yaw Y, and shifted by a vector $\vec{V} = (\eta, \xi, \zeta)$ (rectification transformation) can be described by a homography:

$$H = \begin{pmatrix} p_0 & p_1 & p_2 \\ p_3 & p_4 & p_5 \\ p_6 & p_7 & 1 \end{pmatrix} , \qquad (1)$$

Homography elements are functions of camera attitude parameters:

$$p_0 = U[(-\sin P \sin R \sin Y + \cos R \cos Y)/Z - (\eta/F)(\cos P \sin R)]$$
$$p_1 = U[(\cos P \sin Y)/Z + (\eta/F)\sin P]$$
$$p_2 = U[F(\sin P \cos R \sin Y + \sin R \cos Y)/Z + \eta(\cos P \cos R)]$$
$$p_3 = U[(\sin P \sin R \cos Y + \cos R \sin Y)/Z - (\xi/F)(\cos P \sin R)]$$
$$p_4 = U[(\cos P \cos Y)/Z + (\xi/F)\sin P]$$
$$p_5 = U[F(-\sin P \cos R \cos Y + \sin R \sin Y)/Z + \xi(\cos P \cos R)]$$
$$p_6 = U(-\cos P \sin R)/F$$
$$p_7 = U \sin P / F$$

where $U \equiv \cos P \cos R$; $Z \equiv F/(F+\zeta)$ is a scaling coefficient, and $F \equiv 1/2\tan(FOV/2)$ is a normalized focal length, related to camera field of view FOV.

Consider the simplified case with no roll ($R = 0$), no yaw ($Y = 0$), and a forward-looking camera ($\eta = 0$). Rectification transformation then becomes:

$$\Xi = \begin{pmatrix} 1/Z\cos P & 0 & 0 \\ 0 & 1/Z & -F\tan P/Z \\ 0 & \tan P/F & 1 \end{pmatrix} \qquad (2)$$

Relation between pitch-distorted images I_1 and I_2, and their rectified counterparts I_{1R} and I_{2R} can be written as:

$$I_{1R} = \Xi_1 \bullet I_1 \qquad (3)$$
$$I_{2R} = \Xi_2 \bullet I_2$$

Rectified images are related by translation in ξ:

$$I_{1R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \xi \\ 0 & 0 & 1 \end{pmatrix} \bullet I_{2R} \equiv \mathrm{T}_{12} \bullet I_{2R} \qquad (4)$$

Then the transformation relating pitch-distorted images is:

$$I_1 = \left(\Xi_1^{-1} \mathrm{T}_{12} \Xi_2\right) I_2 = \Omega_{12} I_2 \qquad (5)$$

Assuming constant pitch ($\Xi_1 = \Xi_2$), and denoting $v \equiv 1 + \tan^2 P$, $u_1 = Z\xi$, $u_2 = \tan P / F$, the above transformation can be written as:

$$\Omega_{12} = \begin{pmatrix} \dfrac{v}{v - u_1 u_2} & 0 & 0 \\ 0 & \dfrac{v + u_1 u_2}{v - u_1 u_2} & \dfrac{u_1}{v - u_1 u_2} \\ 0 & \dfrac{-u_1 u_2}{v - u_1 u_2} & 1 \end{pmatrix} \equiv \begin{pmatrix} \omega_0 & 0 & 0 \\ 0 & \omega_4 & \omega_5 \\ 0 & \omega_7 & 1 \end{pmatrix}$$

Elements of transformation Ω_{12} are determined from frames' co-registration process (elements found from this process are denoted as $\hat{\omega}_k$). However these elements contain noise, and a direct calculation of pitch and translation (for example, from equation $\tan P = F\hat{\omega}_7 / (1 - \hat{\omega}_0)$) is extremely inaccurate. To make use of all measured elements, we formulate minimization problem:

$$\min S \equiv \min \sum_{k=0}^{7} (\omega_k - \hat{\omega}_k)^2 \qquad (6)$$

The solution is a point in 8-dimensional ω -space that satisfies theoretical constraints being closest to the point found from registration process of image frames $\hat{\Omega}_{12}$. By using known relationships

$$\omega_4 = 2\omega_0 - 1$$
$$\omega_5 = -(\omega_0 - 1)^2 / \omega_7$$

as constraints, the condition of minimum is reformulated as:

$$\frac{\partial S}{\partial \omega_0} = 2(\omega_0 - \hat{\omega}_0) + 2(\omega_4 - \hat{\omega}_4)\frac{\partial \omega_4}{\partial \omega_0} + 2(\omega_5 - \hat{\omega}_5)\frac{\partial \omega_5}{\partial \omega_0} = 0$$
$$\frac{\partial S}{\partial \omega_7} = 2(\omega_5 - \hat{\omega}_5)\frac{\partial \omega_5}{\partial \omega_7} + 2(\omega_7 - \hat{\omega}_7) = 0$$

or, equivalently

$$(\omega_0 - 1)^4 + \hat{\omega}_5 \omega_7 (\omega_0 - 1)^2 + (\omega_7 - \hat{\omega}_7)\omega_7^3 / 2 = 0$$
$$5(\omega_0 - 1)^2 - (\hat{\omega}_0 - 3 - 2\hat{\omega}_4)(\omega_0 - 1) + \omega_7(\omega_7 - \hat{\omega}_7) = 0$$

Solution of these algebraic equations provides an estimate of elements ω_k based on all data found from the co-registration process, and these elements can be used for calculation of pitch and camera shift.

Similar calculations can be performed when camera tilt has roll component, and camera translation is not restricted to shift in vertical (ξ) direction.

This method has been verified in numerical experiments. The example below uses digital image (Fig.4) as an imaged surface. Acquisition of two frames with camera pitch 30 degrees, from different camera locations separated by 30 pixels was simulated. Acquired frames, with noticeable perspective distortion, are shown in Fig.5.

Fig.4 Digital image used in numerical experiments.

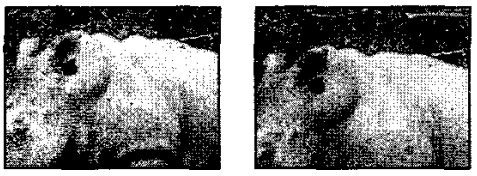

Fig.5. Frames obtained by simulation of acquisition with tilted camera.

Registration of these frames yields the following homography:

$$\hat{\Omega}_{12} = \begin{pmatrix} 0.962119 & 0.000102 & -7.16736e-6 \\ 0.000152 & 0.923645 & -0.065722 \\ -0.000558 & 0.021036 & 1 \end{pmatrix}$$

Minimization (6) results in the following:

$$\Omega_{12} = \begin{pmatrix} 0.961923 & 0 & 0 \\ 0 & 0.923746 & -0.065723 \\ 0 & 0.021041 & 1 \end{pmatrix}$$

Camera pitch calculated from the elements of transformation Ω_{12} was found to be 35.39 degrees, and vertical shift 33.84 pixels. Mosaic of rectified images is presented in Fig.6 and does not show noticeable distortions or inconsistencies.

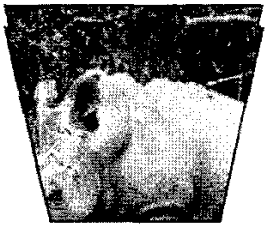

Fig.6. Mosaic created from frames shown in Fig.5, corrected for the pitch-related distortion.

Conclusions

Video collected from deep submergence vehicles is proven to be an invaluable source of information. Despite its lower (in comparison with digital still images) resolution, video footage has the advantage of high overlap between consecutive frames, that allows for successful creation of large area mosaics, even if the camera is tilted, and imaged surface (seafloor) has pronounced 3D content. Video footage mosaicked on board on R/V Atlantis (cruise 11-7, February 2004) after each dive of submersible Alvin, provided geologists and biologists with important information about fault scarps, lava flow morphologies and contacts, and hydrothermal vent communities that could not be accurately determined from looking at separate frames spaced minutes to tens of minutes apart. Mosaics also provided decision-making information for locations of instrument deployments for biological experiments.

An algorithm for estimation of camera tilt from acquired imagery is presented. Numerical experiments indicate that the technique has sufficient accuracy for creation of rectified mosaics for objects' measurements.

Acknowledgments

We thank the pilots and crew of DSV Alvin and the officers and crew of R/V Atlantis, and our colleagues on the AT7-11 cruise in the acquisition of these data. The work was supported by the National Science Foundation – Ocean Sciences Division, Marine Geology & Geophysics Program, and by NOAA grant NA97OG0241. Y.R. thanks Irina Rzhanov for help with the manuscript preparation.

References

[1] D. J. Fornari, "A new deep sea towed digital camera and multi-rock coring system", *Trans. American Geophys. Union, EOS*, vol. 84, 69&73, 2003.

[2] Y. Rzhanov, L. Mayer, D. Fornari, T. Shank, S. Humphris, D. Scheirer, J. Kinsey, L. Whitcomb, "High-Resolution Photo-Mosaicing of the Rosebud Hydrothermal Vent Site and Surrounding Lava Flows, Galapagos Rift 86W: Techniques and Interpretation". *EOS Transactions*, AGU 84 (46), Abstract, OS32A-0231, 2003.

[3] L. G. Brown, "A survey of image registration techniques", *ACM Computing Surveys*, 24, 4, 1992.

[4] Shank, T.M., D. Fornari, D. Yoerger et al. "Deep Submergence Synergy - *Alvin* and *ABE* Explore the Galápagos Rift at 86°W". *EOS, Transactions of the American Geophysical Union* 84: 425-433, 2003.

[5] H. Singh, G. Salgian, R. Eustice, R. Mandelbaum, "Sensor fusion of structure-from-motion, bathymetric 3D, and beacon-based navigation modalities", *International Conference on Robotics and Automation*, Washington DC, USA, 2002.

[6] 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.

[7] Y. Rzhanov, L. M. Linnett, R. Forbes, "Underwater video mosaicking for seabed mapping", *In Proc. ICIP 2002.*

[8] V. N. Vapnik, *Statistical learning theory*, John Wiley & Sons, 1998.

[9] N. Cristianini, J. Shawe-Taylor, *An introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.

[10]M. F. Hutchinson, "On thin plate splines and kriging", *Computing and Science in Statistics*, vol. 25, pp. 55-62, 1993.

[11]F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations", *IEEE Trans. PAMI*, vol. 11(6), pp.567-585, 1989.

[12]K. Zuiderveld, "Contrast limited adaptive histogram equalization", *Graphics Gems IV*, P. Heckbert (ed.), Academic Press, 1994.