Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms

Oscar Pizarro, Ryan Eustice, and Hanumant Singh

Woods Hole Oceanographic Institution, Woods Hole MA 02543, USA,
opizarro@whoi.edu

Abstract. Recent efforts in robust estimation of the two-view relation have focused on uncalibrated cameras with no prior knowledge of pose. However, in practice robotic vehicles that perform image-based navigation and mapping typically do carry a calibrated camera and pose sensors; this additional knowledge is currently not being exploited.
This paper presents three contributions in using vision with instrumented and calibrated platforms. First, we improve the performace of the correspondence stage by using uncertain measurements from egomotion sensors to constrain possible matches. Second, we assume wide-baseline conditions and propose Zernike moments to describe affine invariant features. Third, we robustly estimate the essential matrix with a new 6-point algorithm. Our solution is simpler than the minimal 5-point one and, unlike the linear 6-point solution, does not fail on planar scenes. While the contributions are general, we present structure and motion results from an underwater robotic survey.

1 Introduction

Robust simultaneous estimation of the two view relation [1] has become a key step in most successful structure and motion approaches. Recent efforts have mainly focused on the uncalibrated case with no prior knowledge of pose [2]. However, in practice it is often the case that robotic vehicles carry calibrated cameras as well as pose sensors [3] [4] [5]. In this paper we seek to exploit prior pose knowledge to simplify and improve the reliability of the components used in estimating relative pose from images, and then use this for structure and motion recovery for underwater robotic surveys.

A typical feature-based framework for relative pose estimation is comprised of three main components: correspondence proposal, robust two-view relation estimation with outlier rejection, and final pose refinement. This paper presents three contributions in using this framework with instrumented and calibrated platforms where two view matching is used as the core of a structure and motion estimation algorithm.

First, prior pose knowledge limits the search for correspondences to regions (defined by the motion and its uncertainty). The constrained search increases the reliability of our feature matching stage, which is particularly important when dealing with wide-baseline imagery where inter-image motion may be large. Second, we extend wide-baseline feature descriptors based upon neighborhoods invariant to both affine geometric and affine photometric transformations to incorporate highly discriminating Zernike moments. This improves image matching in the presence of scene structure,

low overlap and significant lighting changes. Finally, we introduce a new 6-point algorithm used within the context of RANSAC to robustly estimate the essential matrix [6] and a consistent set of correspondences. Our six point solution is simpler than the minimal five point solutions [7] [8] and, unlike the linear 6-point solution [9], it does not fail in the presence of planar scenes. This illustrates a significant advantage of the calibrated over the uncalibrated case, noting that in the latter the fundamental matrix cannot be estimated from planar scenes [10].

Results are presented from a robotic underwater vehicle survey, covering an area of approximately $100m^2$ of coral reef in Bermuda. To the best of our knowledge this is the largest 3D reconstruction of an underwater terrain.

1.1 Platform and Approach

Our application is based upon using an autonomous robotic platform for underwater imaging and mapping. We therefore consider a calibrated and pose-instrumented platform to estimate the essential matrix between two image views. Our camera is calibrated using a method based on [11]. The vehicle makes acoustic measurements of both velocity and altitude relative to the bottom. Absolute orientation is measured to within a few degrees using inclinometers and a flux-gate compass, while depth is derived from pressure measurements.

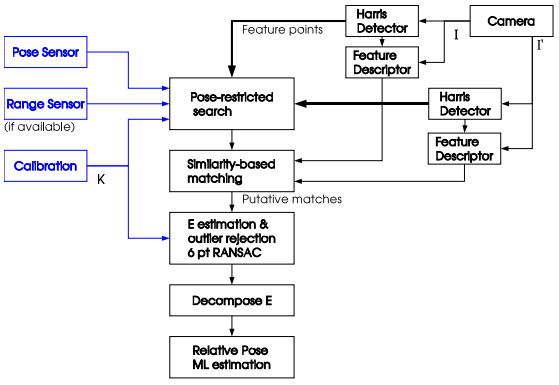


Fig. 1. Overview of our approach to relative pose estimation from instrumented and calibrated platforms. Blue represents additional information compared to the uninstrumented/uncalibrated case. Given two images, we detect features using the Harris interest point detector. For each feature we then determine search regions in the other image by using sensor based pose and depth information. Putative matches are proposed based on similarity and constrained by regions. We then use RANSAC and the proposed 6-point algorithm to robustly estimate the essential matrix which is then decomposed into its proper motion parameters. The pose is then refined by minimizing the reprojection error over all matches considered inliers.

2 Prior Pose Restricted Correspondences

We begin by introducing the two-view point transfer model and then show how to carry forward uncertainty in our pose and calibration to grow this point transfer to a region. This region is used to restrict the interest point matching to a set of candidate correspondences, resulting in increased reliability of similarity-based matching.

2.1 Point Transfer Mapping

Zhang characterized epipolar geometry uncertainty in terms of the covariance matrix of the fundamental matrix [12]. However, that characterization of uncertainty is hard to interpret in terms of physical parameters. Our approach is based on the point transfer mapping which benefits from a more physical interpretation, and an increased robustness by making use of scene range data if available.

In this derivation of the point transfer mapping we assume projective camera matrices $\mathrm{P} = \mathrm{K}[\mathrm{I} \mid \mathbf{0}]$ and $\mathrm{P}' = \mathrm{K}[{}^{c'}_{c}\mathrm{R} \mid {}^{c'}\mathbf{t}_{c'c}]$. Here K is the matrix of intrinsic camera parameters [13], ${}^{c'}_{c}\mathrm{R}$ is the $[3 \times 3]$ orthonormal rotation matrix from camera frame c to c' parameterized by XYZ convention Euler angles $\boldsymbol{\theta}_{cc'} = [\theta_r, \theta_p, \theta_h]^\top$ (roll, pitch, heading) [14], and ${}^{c'}\mathbf{t}_{c'c}$ is the $[3 \times 1]$ translation vector from frame c' to c as represented in frame c'.

Given an interest point with pixel coordinates (u, v) in image I, we define its vector representation $\mathbf{u} = [u, v]^\top$, as well as its normalized homogeneous representation $\underline{\mathbf{u}} = [\mathbf{u}^\top, 1]^\top$. Likewise we define a vector of the imaged scene point as $\mathbf{X} = [X, Y, Z]^\top$ and its normalized homogeneous representation $\underline{\mathbf{X}} = [\mathbf{X}^\top, 1]^\top$. We note that equality in expressions involving homogeneous vectors is implicitly defined up to scale.

The scene point $\mathbf{X}$ is projected through camera P as $\underline{\mathbf{u}} = \mathrm{P}\underline{\mathbf{X}} = \mathrm{K}\mathbf{X}$ which implies that including scale we have

$$\mathbf{X} \equiv Z\mathrm{K}^{-1}\underline{\mathbf{u}} \tag{1}$$

This back-projected scene point can subsequently be reprojected into image I' as

$$\underline{\mathbf{u}}' = \mathrm{P}'\underline{\mathbf{X}} = \mathrm{K}({}^{c'}_{c}\mathrm{R}\mathbf{X} + {}^{c'}\mathbf{t}_{c'c}) \tag{2}$$

By substituting (1) into (2) and recognizing that the following relation is up to scale, we obtain the normalized homogeneous point transfer mapping [13]

$$\underline{\mathbf{u}}' = \mathrm{K}^{c'}_{c}\mathrm{R}\mathrm{K}^{-1}\underline{\mathbf{u}} + \mathrm{K}{}^{c'}\mathbf{t}_{c'c}/Z \tag{3}$$

Finally, by explicitly normalizing the previous expression and defining $\mathrm{H}_\infty = \mathrm{K}^{c'}_{c}\mathrm{R}\mathrm{K}^{-1}$ [13], we recover the non-homogeneous point transfer mapping

$$\mathbf{u}' = \frac{\mathrm{H}_\infty\underline{\mathbf{u}} + \mathrm{K}{}^{c'}\mathbf{t}_{c'c}/Z}{\mathbf{H}^{3T}_\infty\underline{\mathbf{u}} + t_z/Z} \tag{4}$$

where $\mathbf{H}^{3T}_\infty$ refers to the third row of H_∞, and t_z is the third element of ${}^{c'}\mathbf{t}_{c'c}$.

When the depth of the scene point Z is known in camera frame c, then (4) describes the exact two-view point transfer mapping. However, by letting Z vary, (4) describes a functional relationship on Z (i.e. $\underline{\mathbf{u}}' = f(Z)$) which traces out the corresponding epipolar line in I'.

2.2 Point Transfer Mapping with Uncertainty

Typically instrumented pose measurements are made with reference to some fixed world coordinate frame. Then the relative pose measurement required in (4) must be composed by going through this intermediate frame. Expressing the relative pose in terms of measured quantities we have

$$\overset{c'}{c}\mathrm{R} = \overset{w}{c'}\mathrm{R}^{\top}\overset{w}{c}\mathrm{R} \tag{5}$$

$${}^{c'}\mathbf{t}_{c'c} = \overset{w}{c'}\mathrm{R}^{\top}(-{}^{w}\mathbf{t}_{cc'}) \tag{6}$$

To functionally use the point transfer mapping given in (4), we define the 15 element measurement vector $\boldsymbol{\phi}$. The measurement vector $\boldsymbol{\phi}$ is composed from elements of the calibration matrix denoted in vector form as $\mathbf{k}$; the measured pose quantities given in (5) and (6); and the scene depth Z as measured in frame c.

$$\boldsymbol{\phi} = [\mathbf{k}^{\top}, \boldsymbol{\theta}_{cw}^{\top}, \boldsymbol{\theta}_{c'w}^{\top}, {}^{w}\mathbf{t}_{cc'}^{\top}, Z]^{\top} \tag{7}$$

Note that in cases where measurements or constraints can be placed on Z, this information can be used to restrict the mapping from the *entire* epipolar line to a *segment*.

The measured elements of $\boldsymbol{\phi}$ are uncertain. We consider $\boldsymbol{\phi}$ and all quantities derived from it to be random (e.g. $\mathbf{u}'$). Defining $f(\boldsymbol{\phi}; \mathbf{u})$ to be the non-homogeneous point transfer mapping given in (4) and assuming that we know the mean and covariance of the random vector $\boldsymbol{\phi}$ (i.e. $\overline{\boldsymbol{\phi}}$ and $\Sigma_{\boldsymbol{\phi}}$), then to first order we can approximate the mean and covariance of $\mathbf{u}'$ as

$$\overline{\mathbf{u}}' \approx f(\overline{\boldsymbol{\phi}}; \mathbf{u}) \tag{8}$$

$$\Sigma_{\mathbf{u}'} \approx \mathrm{J}\Sigma_{\boldsymbol{\phi}}\mathrm{J}^{\top} \tag{9}$$

where J is the $[2 \times 15]$ Jacobian matrix of $f(\boldsymbol{\phi}; \mathbf{u})$ with respect to $\boldsymbol{\phi}$ evaluated at $\overline{\boldsymbol{\phi}}$.

Figure 2 illustrates the search regions for an image pair.

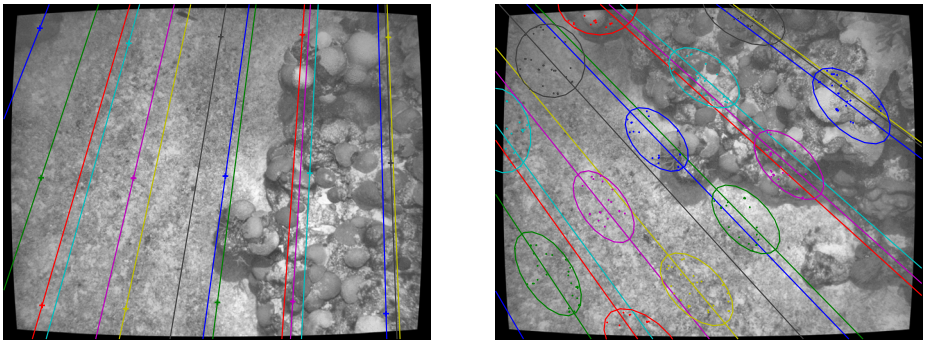
3 Feature Extraction, Description and Matching

We relate images using a feature-based approach under wide-baseline imaging conditions with changing illumination and unknown scene structure. A modified Harris corner detector [15] extracts interest points by selecting local maxima of the smaller eigenvalue of the second moment matrix.

3.1 Orientation normalization

Attitude sensors provide information on the 3D orientation for cameras c and c' in a fixed reference frame w. This allows for normalization of orientation viewpoint effects via a homography. The infinite homography, $\mathrm{H}_{\infty} = \mathrm{K}_{a}^{b}\mathrm{R}\mathrm{K}^{-1}$ warps an image taken from camera orientation a into an image taken from camera orientation b.

Given 3D camera rotation matrices $\overset{w}{c}\mathrm{R}$ and $\overset{w}{c'}\mathrm{R}$ generated from vehicle orientation measurements, we can warp images I and I' each into a canonical viewpoint coordinate frame oriented parallel with frame w (e.g. the warped images correspond to a camera coordinate frame oriented with North, East, Down).



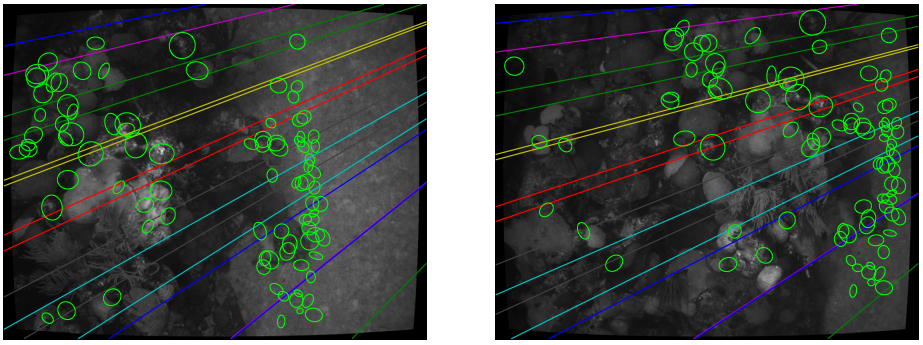
(a) (b)

Fig. 2. Prior pose restricted correspondence search on a pair of underwater coral reef images. (a) A sampling of interest points are shown in the left image along with their color coded sensor instantiated epipolar lines. (b) The right image shows the corresponding color coded constrained search regions for the interest points in the left image (95% confidence level); the sensor instantiated epipolar lines; and the candidate interests points which fall within these regions. The search regions are determined using an altimeter measurement of the average scene depth and setting σ_Z to 0.75 meters

3.2 Neighborhood Extraction and Description

We determine a neighborhood around each interest point that is invariant to affine geometric transformations using a modified version of the method proposed by Tuytelaars [16]. In essence, we sample a 31×31 pixel neighborhood along lines radiating from the interest point. For each line we select the extrema of an affine invariant function (maximum difference in intensities between the interest point and points along the ray). The convex hull of these maximal points is approximated with an elliptical neighborhood which is then mapped onto the unit circle. To increase discriminating power, a second neighborhood twice as large as the first is also mapped onto a unit circle. These circular patches are normalized for affine photometric invariance (demeaned and normalized by their energy content).

Moment-based descriptors [17] have shown promise in describing image regions for matching purposes. We chose to use Zernike moments for this purpose as they are compact (generated from an orthogonal complex polynomials) and highly discriminating [18]. Typically only the magnitude of Zernike moments are used since this provides rotational invariance, but having pre-compensated for orientation using attitude sensors we use the full complex moments. The pair of circular patches are each described by the first 25 complex Zernike coefficients, making the total feature descriptor a vector of 50 complex numbers for the two areas. Similarity is measured as the real part of the product of the feature descriptors. To improve inter-image discrimination, features that are similar to other features intra-image are considered unreliable and removed from the matching process.



(a) (b)

Fig. 3. Example of putative matches obtained on typical underwater imagery by only using similarity measures (i.e. not restricted by pose prior to illustrate capability). Each affine invariant neighborhood is shown at the larger of the scales as a green ellipse. Note that most matches are correct. Approximate epipolar lines from navigation are also shown.

4 Essential Matrix Estimation

Relative pose from calibrated cameras is a 5 DOF problem (3 DOF for rotation and 2 DOF for direction of motion between cameras), because of loss of scale. Minimal 5-point algorithms [7][8][19] tend to be ill-posed, of complex implementation, and can present up to 20 solutions that then have to be tested. In this section we investigate the possibility of using 6 point matches to determine relative motion using the essential matrix.

The $[3 \times 3]$ essential matrix E encodes the relative motion between two cameras [13]. In term of the motion parameters $\mathrm{E} = [{}^{c}\mathbf{t}_{c'c}]_{\times}\,{}^{c'}_{c}\mathrm{R}$. The essential matrix can be considered as a special case of the fundamental matrix, satisfying the following relationship:

$$\underline{\mathbf{x}}'^{\top}\mathrm{E}\underline{\mathbf{x}} = 0 \tag{10}$$

where $\underline{\mathbf{x}} = [x, y, 1]^{\top}$ and $\underline{\mathbf{x}}' = [x', y', 1]^{\top}$ are normalized image point correspondences (i.e. $\underline{\mathbf{x}} = \mathrm{K}^{-1}\underline{\mathbf{u}}$ and $\underline{\mathbf{x}}' = \mathrm{K}^{-1}\underline{\mathbf{u}}'$). As a fundamental matrix, E has a null determinant and because of calibration it has two equal singular values. Define

$$\mathbf{e} = [e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}]^{\top} \tag{11}$$

as the vector representation of (E). Then (10) can be written as

$$\left[\, x'x \;\; x'y \;\; x' \;\; y'x \;\; y'y \;\; y' \;\; x \;\; y \;\; 1 \,\right] \mathbf{e} = 0 \tag{12}$$

With a set of n point matches we can form a linear system of the form

$$\mathrm{A}\mathbf{e} = \mathbf{0}_{n\times 1} \tag{13}$$

where each row of A has the form of the row vector in (12).

For $n = 8$ and non-critical motion and point configurations, we have the classic 8-point algorithm [20] which solves for the $\mathbf{e}$ that satisfies (13) . With $n = 7$ we can find $\mathbf{e}$ as the linear combination of the two generators of the right null space of A and impose the $det(\mathrm{E}) = 0$ constraint. However for points in a plane the rank of A drops to 6, and the 8 and 7-point algorithms can no longer be used [9]. With $n = 6$ point matches, A will have rank 6 and $\mathbf{e}$ will be a linear combination of the generators of the right null space of A determined by SVD, i.e. $\mathbf{e} = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3$. Homogeneity of the equations implies that $\mathbf{e}$ is only determined up to scale, and therefore can be expressed in terms of two parameters $\mathbf{e} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 + \mathbf{e}_3$ or in matrix form

$$\mathrm{E} = \alpha\mathrm{E}_1 + \beta\mathrm{E}_2 + \mathrm{E}_3 \; . \tag{14}$$

The values of α and β must be determined such that E is an essential matrix. A $[3 \times 3]$ matrix is an essential matrix (one null and two equal singular values) if and only if it satisfies the Demazure constraint [6]

$$\mathrm{E}\mathrm{E}^\top\mathrm{E} - \frac{1}{2}trace(\mathrm{E}\mathrm{E}^\top)\mathrm{E} = 0 \tag{15}$$

By replacing E in (15) with (14) we generate a system of 9 homogeneous polynomial equations of degree 3 in α and β. This can be considered a homogeneous linear system in the monomials of α and β: α^3, $\alpha^2\beta$,α^2,$\alpha\beta^2$, $\alpha\beta$, α, β^3, β^2, β,1. With 9 linear equations and 9 unknowns it is possible to solve uniquely for the vector of unknowns and therefore obtain E. This technique is known as the 6-point linear algorithm [21] and will fail in cases where all of the 3D points are configured in a plane. In such a case system A will still have rank 6, but the system defined by (15) will now drop to rank 4 rather than 9, because the linear dependence between 3D points introduces additional relations [9]. This result significantly reduces the applicability of the linear 6-point algorithm in practical situations. It could be used in a model selection framework [10], but we seek a simpler approach.

4.1 A 6-Point Algorithm Robust to Planar Scenes

Our method uses only 4 equations from the Demazure constraint and solve the resulting system of polynomials of degree 3. By always using a system of rank 4 we will find a solution even in the presence of planar scenes. We show that by manipulating this system we can generate a polynomial of degree 6 in β and then solve for α.

Consider four equations from the Demazure constraint in terms of the monomials in α and β. To pick four equations, we perform SVD on the 9 equations of the Demazure constraint and then select the four right singular vectors associated with the largest singular values. Performing Gauss-Seidel elimination on the four equations we have

$$\begin{array}{cccccccccc}
\alpha^3 & \alpha^2\beta & \alpha^2 & \alpha\beta^2 & \alpha\beta & \alpha & \beta^3 & \beta^2 & \beta & 1 \\
\hline
1 & & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & 1 & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & & 1 & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & & & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array} \tag{16}$$

Here a blank represents zeros and '$\cdot$' represent some value not eliminated by Gauss-Seidel. This system can be represented as

$$\alpha^3[0a] + \alpha[1b] + [3c] = 0 \tag{17}$$
$$\alpha^2[1d] + \alpha[1e] + [3f] = 0 \tag{18}$$
$$\alpha^2[0g] + \alpha[1h] + [3i] = 0 \tag{19}$$
$$\alpha[2j] + [3k] = 0 \tag{20}$$

where the number in brackets represents the degree of a polynomial in β and the letter is an identifier for that particular polynomial. By multiplying the second equation of the system (18) by $[0g]$ and the third equation (19) by $[1d]$ and then subtracting we obtain

$$\alpha([1e][0g] - [1h][1d]) + [3f][0g] - [3i][1d] = 0\,. \tag{21}$$

Notice that the above expression no longer depends on α^2. Defining $[2s] = [1e][0g] - [1h][1d]$ and $[4t] = [3f][0g] - [3i][1d]$, together with the fourth equation (20), we have

$$\alpha[2s] + [4t] = 0 \tag{22}$$
$$\alpha[2j] + [3k] = 0 \tag{23}$$

Cross multiplying by the polynomials of degree two and subtracting we obtain a single polynomial equation of degree six

$$[4t][2j] - [3k][2s] = 0 \tag{24}$$

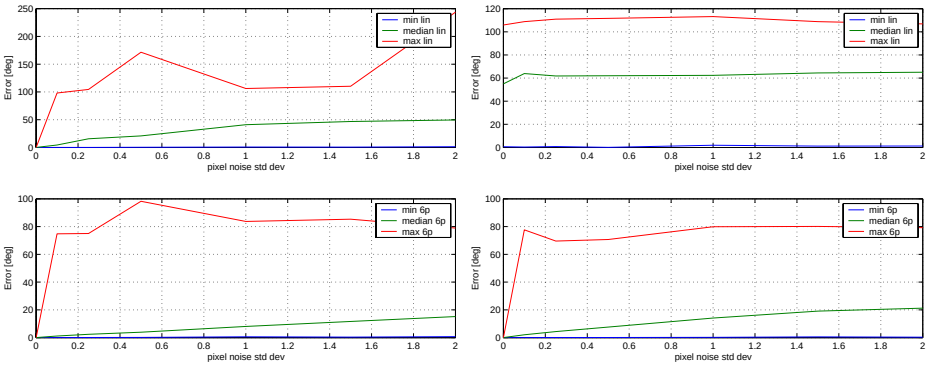
We solve this polynomial and use the real roots to then solve for α using (20). For each pair (α, β) that is real, we calculate the corresponding essential matrix according to (14). The proposed 6-point algorithm will produce up to six possibly valid essential matrices.

Using synthetic data sets (generated for both planar and non-planar scenes) and random relative motion, we have determined that one of the essential matrices produced by this 6-point algorithm always corresponds to the true camera motion for perfect (noise free) measurements.

4.2 Robust Estimation of the Essential Matrix

The reliability of this algorithm in the presence of noisy measurements is still under investigation. To test how the performance of this algorithm degrades in the presence of noise, we performed 1000 trials with randomly generated scenes and motions. For each trial the essential matrices computed by the 6-point algorithm were decomposed into their respective rotation and translation representation. The essential matrix with rotation and translation that was closest (minimum error) to the true motion was selected. In order to summarize results in one quantity, we defined a pose error measure as the sum of (1) the angle of rotation between the true rotation matrix R and the estimated $\hat{\mathrm{R}}$ using the axis-angle representation [22], and (2) the angle between the translation direction vectors. These trials were then repeated for different levels of noise added to the pixel coordinates.

Figure 4 shows the evolution of the error as noise increases for points in general and planar configurations. Notice that for general configurations and perfect measurements (i.e. zero noise) both algorithms produce the correct essential matrix but that for planar scenes the linear 6-point algorithm fails even for perfect measurements.



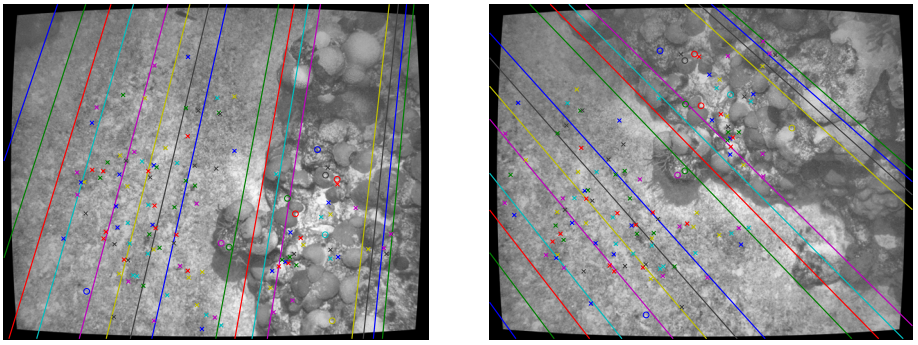
(a) (b)

Fig. 4. Noise test with general scenes. The minimum, median, and maximum pose error (defined as the sum of the rotation error and the angular error of the baseline direction vector) over 1000 trials, plotted against noise variance. (top) Linear 6-point algorithm, (bottom) proposed 6-point algorithm. Noise test with planar scenes. The minimum, median, and maximum pose error (defined as the sum of the rotation error and the angular error of the baseline direction vector) over 1000 trials, plotted against noise variance. (top) Linear 6-point algorithm, (bottom) proposed 6-point algorithm. Notice the failure of the linear 6-point algorithm even in the absence of noise.

Even though the proposed 6-point algorithm degrades in the presence on noise, Figure 4 shows that a large number of estimates will be close to the true motion. This suggests that the algorithm can be used effectively in a RANSAC context where it is reasonable to expect there will be point combinations that will yield an essential matrix close to the true one and will explain a large fraction of the correctly matched points. Our results with real data from an underwater vehicle have been excellent. Figure 5 shows the resulting image-based points considered inliers by RANSAC. The epipolar geometry in the figure is a refinement by maximum likelihood estimation (MLE) of the RANSAC estimate.

5 Application to Structure and Motion

The robust two view matching for instrumented calibrated platforms is at the center of a system currently being developed for structure and motion recovery from underwater robotic surveys. Most of the processing involves establishing relative poses between cameras. This is done without reference to a global coordinate system and has



(a) (b)

Fig. 5. Epipolar geometry and correspondences. The given image pair illustrates the maximum likelihood refined image-based epipolar geometry. RANSAC determined 106 consistent inliers designated 'x', from the putative set of 116 matches. The rejected outliers are designated 'o'.

three attractive characteristics: only requires calculating globally consistent poses before bundle adjustment, avoids local minima in the structure solution, does not use local structure and is therefore less sensitive to local mismatches. Our approach to reconstruction is similar in spirit to the Atlas framework [23], except that our images only provide direction to features but not range, and that our poses have 6DOF given that the camera is mounted on an underwater vehicle.

5.1 Results

Figure 6 shows results from an underwater robotic survey of a section of coral reef in Bermuda. Bottom depth was approximately $21m$, shallow enough to be under the effect of surface waves (notice that ripple patterns formed on the sandy bottom).

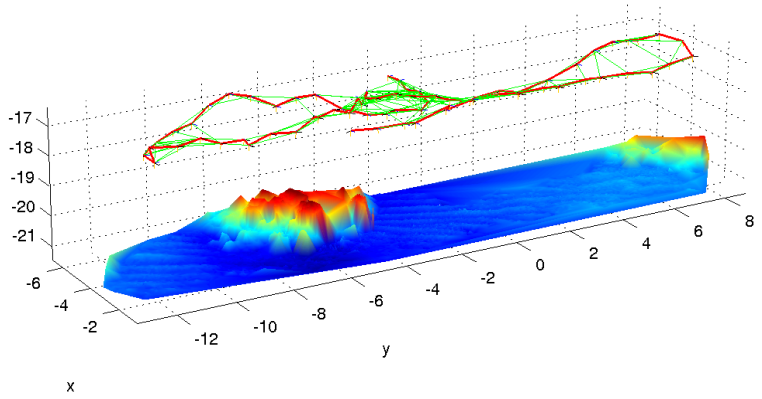
6 Conclusions and Future Work

We have presented a a two-view matching approach applicable to instrumented and calibrated imaging platforms. It has succesfully been used in wide-baseline, large scale reconstructions of underwater scenes from imagery acquired by a robotic vehicle. Future work will extended and refine the structure and motion recovery to hundreds or thousands of images.

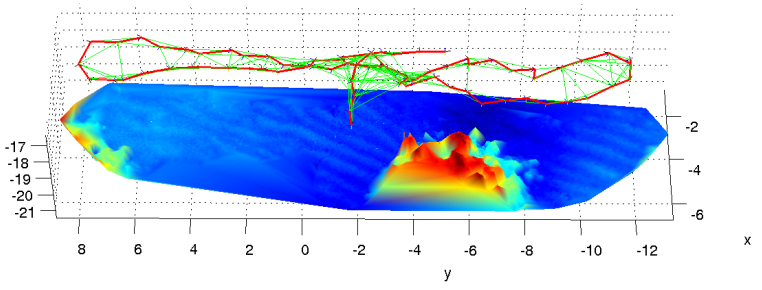
References

1. M. A. Fischler and R Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

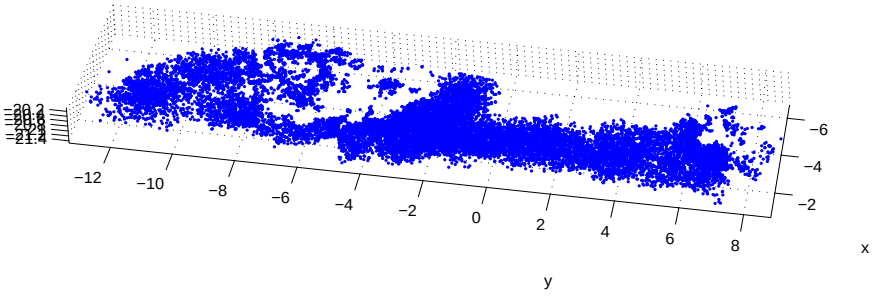
2. M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Metric 3d surface reconstruction from uncalibrated image sequences. In *3D Structure from Multiple Images of Large Scale Environments, LNCS Series*, volume 1506, pages 139–154. Springer-Verlag, 1997.
3. M. Bosse. *A Vision Augmented Navigation System for an Autonomous Helicopter*. Masters, Boston University, Date 1997.
4. M. Antone and S. Teller. Scalable extrinsic calibration of omni-directional image networks. *International Journal of Computer Vision*, 49(2-3):143–174, 2002.
5. A.J. Davison. *Mobile Robot Navigation Using Active Vision*. Ph.d. thesis, University of Oxford, Date 1999.
6. O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.
7. B.K.P Horn. Relative orientation. *International Journal of Computer Vision*, 4(1):58–78, 1990.
8. B. Triggs. Routines for relative pose of two calibrated cameras from 5 points. Technical report, INRIA, 2000.
9. J. Philip. Critical point configurations of the 5-,6-,7-, and 8-point algorithms for relative orientation. Technical Report TRITA-MAT-1998-MA-13, KTH Royal Institute of Technology, 1998.
10. P. Torr, A. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated images. *International Journal of Computer Vision*, 32(1):27–44, August 1999.
11. Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.
12. Z. Zhang. On the optimization criteria used in two-view motion analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7):717–729, 1998.
13. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
14. T.I. Fossen. *Guidance and Control of Ocean Vehicles*. John Wiley and Sons Ltd., New York, Date 1994.
15. C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Conference*, pages 189–192, Manchester, UK, August 1988.
16. T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proceedings of the British Machine Vision Conference 2000*, pages 736–739, Bristol, UK, 2000.
17. F. Mindru, T. Moons, and L. Van Gool. Recognizing color patterns irrespective of viewpoint and illumination. In *Proceedings of CVPR99*, pages 368–373, 1999.
18. W.Y. Kim and Y.S. Kim. A region-based shape descriptor using Zernike moments. *Signal Processing:Image Communication*, 16(1-2):95–102, September 2000.
19. O. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4(3):225–246, 1990.
20. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
21. J. Philip. A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. *Photogrammetric Record*, 15(88):589–599, October 1996.
22. X. Pennec and J-P. Thirion. A framework for uncertainty and validation of 3d registration methods based on points and frames. *International Journal of Computer Vision*, 25(3):203–229, 1997.
23. M. Bosse, J. Newman, P.and Leonard, and S. Teller. An atlas framework for scalable mapping. Technical Report MIT Marine Robotics Laboratory Technical memorandum 2002-04, 2002.



(a)



(b)



(c)

Fig. 6. Example of structure and motion recovery from a section of 71 images of an underwater robotic survey. The images (a) and (b) represent two views of the reconstruction as a surface using a Delaunay triangulation. They also contain the camera positions - red lines highlight the sequence in which the images were taken while the green lines represent additional overlapping imagery (related by our two-view algorithm). The bottom figure represents the 3D reconstruction of over 19000 feature points. Units are in meters relative to the start of the survey in (x,y) and to sea surface in z. The reconstruction corresponds to an area of approximately $100m^2$.