Trajectory Reconstruction with Uncertainty Estimation using Mosaic Registration

Nuno Gracias, José Santos-Victor

Instituto Superior Técnico & Instituto de Sistemas e Robótica, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

Abstract

This paper addresses the problem of estimating the 3-D trajectory and associated uncertainty of an underwater autonomous vehicle from a set of images of the seabed taken by an onboard camera. The presented algorithms resort to the use of video mosaics and build upon previous work on image registration and visual pose estimation. The pose estimation is accomplished in two steps. Firstly a video mosaic is created automatically, covering a region of interest of the seabed. Then, after associating a 3D referential for the mosaic, the estimation of the camera position from a new view of the scene becomes possible.

The main contribution of this paper lies on the assessment of the performance of the 3D pose algorithms. In order to do this, an image sequence with available ground-truth is used for precise error measuring. A first order error propagation analysis is presented, relating the uncertainty in the location of the match points with the uncertainty in the pose parameters. The importance of predicting the estimate uncertainty is emphasized by the fact that it can be used for comparing algorithms and for the on-line monitoring of the vehicle trajectory reconstruction quality.

Several iterative and non-iterative pose estimation methods are discussed, differing both on the criteria being minimized and on the required information about the camera intrinsic parameters. This information ranges from the full knowledge of the parameters, to the case where they are estimated using self-calibration from an image sequence under pure rotation.

The implemented pose algorithms are compared for the accuracy and estimate covariance.

Keywords: Underwater computer vision; video mosaics; trajectory reconstruction, uncertainty estimation

1. Introduction

In the last few years, computer vision has increasingly been used as a sensing modality for underwater vehicles used in tasks where accurate measures at short range are needed [1]. A considerable amount of research interest has been directed towards providing autonomy to underwater vehicles using vision, namely in self-location and motion estimation.

The work described in this paper addresses the issues of vehicle self-location and uncertainty estimation using video mosaics as visual maps. Having such maps referenced to a world coordinate system, enables a camera-equipped autonomous vehicle (in an unknown position and orientation) to locate itself once it has found the correct mapping from the mosaic to the image frame. The approach for automatic creation of video mosaics builds upon our previous work[2] and is based on image motion estimation in a robust and automatic way. We deal with the issue of determining the 3D position and orientation of a vehicle from new views of a previously created mosaic. The problem of pose estimation is addressed, using the available information on the camera intrinsic parameters. This information ranges from the

*E-mail: ngracias,jasv@isr.ist.utl.pt

full knowledge, to the case where they are estimated using a self-calibration technique, based on the analysis of an image sequence captured under pure rotation. Direct pose estimation methods are presented based on the initial computation of the image-to-world homography, which are suitable for real-time operations on setups of limited computing power. These methods are further refined using iterative optimization, applied to the minimization of explicit error functions both on the matched coordinates space or on the elements of the image-to-world homography. One of the benefits of using such error functions lies on the fact that it enables the analysis of uncertainty propagation, using a first order Taylor series approximation of the mapping between observations and estimated pose parameters.

The importance of the uncertainty propagation prediction is threefold. Firstly, it provides quantitative measures for the comparison of different pose estimation algorithms. Secondly, it allows the detection of eventual degenerate parameter configurations. Thirdly, for practical setups, it allows the on-line monitoring of the quality of trajectory reconstruction. This last aspect is of utmost importance in situations where the risk of loosing a vehicle, due to poor positioning, bears very high costs.

Trajectory recovery results are presented for an image sequence for which ground-truth is available. The presented techniques are suitable for autonomous underwater vehicle (AUV) navigation near a flat oceanic floor, where a planar map is an accurate representation of the environment. A possible application scenario for these methods is in underwater archeological site exploration or in marine geological surveys, where an AUV is required to do an initial area mapping followed by periodic inspections.

A method for 3D motion estimation and mosaic construction was proposed by Xu *et al.* [3] and tested on a floating platform. The use of mosaics as a tool to provide visual maps for navigation has been explored by Zheng *et al.*[4], in the context of land robotics and route recognition. In their work, a visual memory of the motion of a mobile robot is created in the form of panoramic mosaics that are later used for robot positioning. However, the visual representations are used solely for navigation purposes and the panoramic views created do not correspond to geometrically and visually correct mosaics. Over the years, the problem of camera pose estimation has been thoroughly addressed in the Computer Vision literature. For recent progress in linear methods in pose estimation refer to [5] and the references there in. The direct, non-iterative, pose estimation algorithms presented in this paper decompose an image-to-mosaic homography matrix, in order to find the rotation matrix and displacement vector relating the camera frame to a world frame (extrinsic parameters). In this sense, it relates to the work by Ganapathy[6], where the extrinsic parameters are recovered directly from a camera projection matrix.

The paper is organized as follows. Section 2 describes the camera model decomposition, the process of creating video mosaics, and the notation required for the methods described later. Section 3 is devoted to the registration of new views on a previously constructed mosaic and to the problem of estimating the camera pose. Finally, Section 6 summarizes and draws some conclusions on the performance and applicability of the methods.

2. Geometric background

2.1. Camera Model

For the purposes of this paper, an useful decomposition of the standard (3×4) camera matrix P is $P = K\,[{}^{C}_{W}R \quad {}^{C}_{W}\mathbf{t}]$, where K is the (3×3) upper-diagonal intrinsic parameter matrix, ${}^{C}_{W}R$ is the rotation matrix relating the orientation of the 3D camera and world frame, and ${}^{C}_{W}\mathbf{t}$ is the location of the world origin in camera frame coordinates. The intrinsic parameter matrix has the form

$$K = \begin{bmatrix} fk_u & fk_\theta & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where k_u and k_v are scaling factors (along u and v), and (u_0, v_0) is the location of the principal point and f is the focal length. The additional parameter k_θ gives the skew between axes. For the most commonly used CCD cameras, k_θ can be considered zero on applications not relying on

highly accurate calibration. The intrinsic parameter matrix can be estimated from P by means of the QR factorization[7].

2.2. Mosaic Creation

The mosaicing techniques used in this paper are detailed in [7], and are summarized in this section.

The mosaic creation evolves from the analysis of point correspondences in order to estimate the homographies between pairs of consecutive frames of the video sequence. For each pair of images, a set of points is selected from one image and matched over the other, minimizing the sum of squared differences of the pixel intensities of the neighboring areas around the points. Due to the error prone nature of this matching process, it is expectable that a number of point correspondences will not relate to the same 3-D point. Therefore, a robust selection of the matched points is essential for the accurate image motion estimation. For this, a random sampling algorithm was implemented[8], using a minimization criteria based on the median of the error distances between the point projections and their expected locations using the estimated homography.

After estimating the frame-to-frame homographies, these are cascaded to form a global registration, where all frames are mapped into a common, arbitrarily chosen, reference frame. For the purposes of this paper, the reference frame is computed using some manually selected world points with known metric coordinates.

The following step consists in merging the images. On overlapping regions, some method has to be established in order to determine the unique intensity value that will be used on the final mosaic. The most commonly used is the median operator, which is adequate for underwater sequences of the seabed where moving fish or algae are captured. Figure 1 presents a mosaic from a sequence of images captured by a surface-driven ROV, on a pipeline inspection task. Although the original sequence presents noticeable perspective distortion effects, a reference frame was chosen as to make the contour lines of the pipeline approximately parallel, yielding a top view of the floor.

3. Pose Estimation from Planar Scenes

We will now describe two sets of methods for the pose estimation, which differ on the minimizing criteria and on the required intrinsic parameter information.

The first set comprises non-iterative methods which provide fast pose estimates, adequate for real-time applications on vehicles with limited computational capabilities. These methods are based on the initial estimation of a image-to-world homography $T_{image,World}$.

The methods of the second set refine the estimates of the first, by using iterative optimization procedures to minimize explicit cost functions. These methods have considerable higher computational requirements, but allow a first order analysis of the error propagation and the computation of the pose parameters uncertainty.

The $T_{image,World}$ homography can be obtained by cascading an image-to-mosaic homography $T_{image,mosaic}$ (computed using the techniques described above for the mosaic creation) with a $T_{mosaic,World}$ homography, that relates the mosaic image frame with their metric counterparts on the world. In this paper we do not address the problem of finding the appropriate $T_{mosaic,World}$, but take into account the effect of both its nominal value and uncertainty, in the form of a covariance matrix of its elements. Also we assume, without loss of generality, that the world frame is such that all the points in the planar scene have null $\overrightarrow{z}$coordinate.

3.1. Non-iterative Methods

3.1.1. Known intrinsic parameter matrix

For the case where the intrinsic parameter matrix is known, a simple and useful decomposition can be obtained for the homography $T_{image,World}$ which relates planar world points with their camera projections. Let L be a (3×3) matrix containing the first two columns of ${}^{C}_{W}R$, and the vector, ${}^{C}_{W}\mathbf{t}$. Then $T_{image,World}$ can be decomposed as

$$T_{image,World} = \lambda K L \tag{1}$$

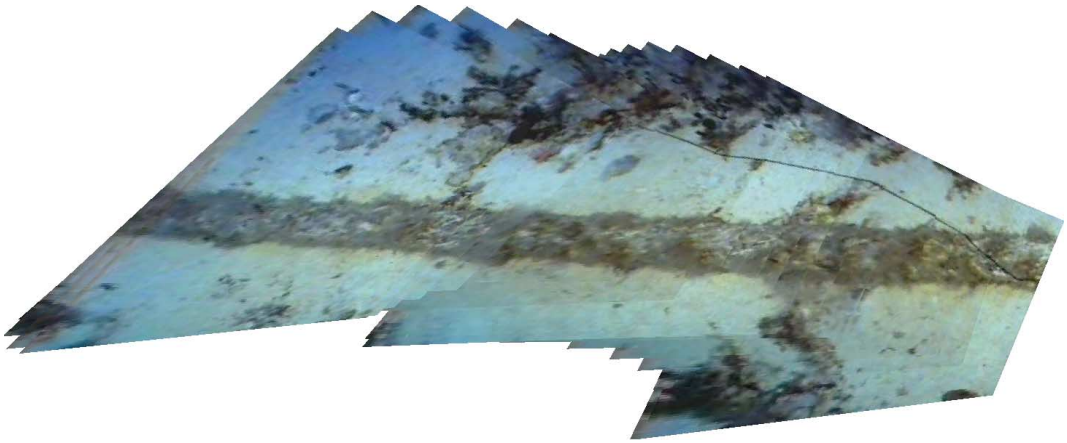


Figure 1. Underwater pipeline mosaic example. A useful reference frame was chosen yielding a top view of the sea floor.

where λ is an unknown scale factor. In order to recover the pose information embedded in L, one has first to estimate the unknown scale factor λ. By taking into account the fact that the first two columns of L have unit norm, a constraint can be imposed on $\|\lambda\|$. A straightforward estimator for $\|\lambda\|$ is the average of the norms of the first two columns of $\lambda L = K^{-1}T_{image,World}$.

For the recovery of ${}^{C}_{W}R$, one has to impose orthogonality on the first two columns, and compute their cross product to obtain the last column. Let ${}^{C}_{W}R_1$ and ${}^{C}_{W}R_2$ be the two candidates for ${}^{C}_{W}R$, corresponding respectively to the scaling by $+\|\lambda\|$ and $-\|\lambda\|$. The matrices ${}^{C}_{W}R_1$ and ${}^{C}_{W}R_2$ relate by ${}^{C}_{W}R_1 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} {}^{C}_{W}R_2$

The corresponding optical centre locations are given by

$$ {}^{W}_{C}\mathbf{t}_1 = -\tfrac{1}{\|\lambda\|}\, {}^{C}_{W}R_1^T \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} $$

$$ \text{and } {}^{W}_{C}\mathbf{t}_2 = \tfrac{1}{\|\lambda\|}\, {}^{C}_{W}R_2^T \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} $$

where $\begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^T$ is the last column of λL. The locations of the optical centres differ by the last coordinate which is symmetric. Both solutions for ${}^{C}_{W}R$ and ${}^{W}_{C}\mathbf{t}$ are in accordance with $T_{image,World}$, and are geometrically valid. In the application of this work, we are only interested in the positive $\overrightarrow{z}$ axis solution for ${}^{W}_{C}\mathbf{t}$, which corresponds to the camera being above the plane of the floor.

3.1.2. Known principal point and skewing

An alternative method for estimating the camera pose can be devised if only the principal point location and the skewing ratio $\frac{fk_\theta}{fk_v}$ are known, instead of the full K matrix. Let us decompose K as the product of an upper triangular matrix U, with ones on its diagonal, by a diagonal matrix A, such that

$$ K \doteq UA = \begin{bmatrix} 1 & \frac{fk_\theta}{fk_v} & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} fk_u & 0 & 0 \\ 0 & fk_v & 0 \\ 0 & 0 & 1 \end{bmatrix} $$

Since U is invertible, one can extend Equation (1) to

$$ U^{-1}T_{image,World} \doteq AL \qquad (2) $$

The left side of Equation (2) can be computed from image measurements. As we are interested

in estimating the unknown intrinsic parameters in the A matrix, we will start by explicitly including an unknown scale factor λ, in order to remove the equality up to scale. Let $M = U^{-1}T_{image,World}$. Equation (2) can thus be written as $A(\lambda L) = M$. By considering the first two columns of this equality, where the unknown scale factor λ has been multiplied to the elements of L, and by imposing the additional conditions of equal norm and vector orthogonality, a system of equations on fk_u and fk_v can be written in the form of

$$\begin{bmatrix} m_{11} \cdot m_{12} & m_{21} \cdot m_{22} \\ m_{11}^2 - m_{12}^2 & m_{21}^2 - m_{22}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{(fk_u)^2} \\ \frac{1}{(fk_v)^2} \end{bmatrix}$$

$$= - \begin{bmatrix} m_{31} \cdot m_{32} \\ m_{31}^2 - m_{32}^2 \end{bmatrix}$$

After estimating fk_u and fk_v the pose can be recovered using the method described above, for known K matrix. For the experimental part of this work, the skew was assumed to be zero.

3.1.3. Self-calibration from a rotating camera

A self-calibration method can be used for the estimation of the K matrix, if a sequence of images taken by a camera with constant intrinsic parameters and undergoing pure rotation, is available. This method does not require any knowledge on the scene structure, nor the rotation of the camera frame between images. Therefore, it is specially suited to applications where on-line calibration is required and the camera can be rotated around its optical center.

The theory behind this method is presented by Hartley in [9]. For the case of stationary cameras (where no translation is allowed), the mapping between corresponding points in two views is represented by a 2D homography $T_{i,j} = KR_{j,i}K^{-1}$. The homography can be computed directly from image measurements, and depends only on the intrinsic parameter matrix and on the camera rotation $R_{j,i}$ between the two images. As noted in [9], $T_{j,i}$ is only meaningfully defined up to scale, but taking into account the fact that the product $KR_{j,i}K^{-1}$ has unit determinant, the exact equality $T_{i,j} = KR_{ji}K^{-1}$ will hold if $T_{i,j}$ is scaled by an appropriate factor.

A linear system of equations, not depending on the rotation matrices, can be constructed on the elements of the symmetrical matrix $C = KK^T$, and solved using the SVD [9]. The recovery of K can be achieved if C is positive-definite, by means of the Choleski decomposition[10] and is unique if K is assumed to have positive diagonal entries. For noise-free data, C is positive-definite by construction, and for noisy data it might not be so.

3.2. Iterative Methods

For the iterative estimation of the image-to-mosaic homography $T_{i,mosaic}$, we define the following scalar error function

$$F(X, \Theta) = \sum_{n=1}^{N} [d^2 \left(x_n, T_{i,mosaic} \cdot x'_n\right)$$

$$+d^2 \left(x'_n, T_{i,mosaic}^{-1} \cdot x_n\right)]$$

where $d\left(x_n, T_{i,mosaic} \cdot x'_n\right)$ is the euclidean distance between the point x_n and the projection of the corresponding point of the mosaic x'_n. The coordinate vector X contains the coordinates of the matched points, and Θ is the estimate vector containing a parameterization of the elements of $T_{i,mosaic}$. The use of the two distance terms has to do with the fact that the point projections, on the image and on the mosaic, do not play a symmetric role. By using the two terms under the assumption of independent Gaussian noise on the coordinates, the residuals of this criteria are approximately Gaussian thus yielding an approximation to a maximum likelihood estimate.

In order to estimate Θ, $F(X, \Theta)$ needs to be minimized using the coordinates of at least 4 matched points. We have chosen a parameterization Θ, allowing unconstrained minimization, in which the first 8 elements of $T_{i,mosaic}$ are considered, after normalizing by dividing by $T_{i,mosaic}(3,3)$. This parameterization does not represent an homography that maps the origin of the mosaic image frame onto the infinity in the other image frame, but this condition was not found to be of practical importance.

As a starting point for the minimization of

$F(X,\Theta)$, the result of the least squares computation of the homography is used, for the initial value of the parameter vector.

3.2.1. Pose from the Image-to-World Homography

For the case of pose estimation, the used error function is

$$F(X,\Theta) = \|T_{i,World} - \Psi(K,\Theta)\|_{Frob}$$

where $\Psi(K,\Theta)$ is an image-to-world homography constructed using the pose parameters Θ and the camera intrinsics. The pose parameters are represented by the 6-vector $\Theta = \begin{bmatrix} \alpha & \beta & \gamma & {}^{W}_{C}\mathbf{t}_x & {}^{W}_{C}\mathbf{t}_y & {}^{W}_{C}\mathbf{t}_z \end{bmatrix}$ containing the 3 camera rotation angles and the location of the camera centre in world coordinates. For this calculation the input data X consists of the first 8 elements of the normalized $T_{i,World}$.

When only the location of the principal point and skew is known, instead of the complete intrinsic parameter matrix, the above error function can be used with minor modifications. In this case

$$F(X,\Theta) = \|T_{i,World} - \Psi(u_0, v_0, fk_\theta, \Theta)\|_{Frob}$$

and the parameter vector also contains the unknown intrinsics,

$$\Theta = \begin{bmatrix} \alpha & \beta & \gamma & {}^{W}_{C}\mathbf{t}_x & {}^{W}_{C}\mathbf{t}_y & {}^{W}_{C}\mathbf{t}_z & fk_u & fk_v \end{bmatrix}$$

.

3.2.2. Pose from Matched Points

The use of an optimization procedure for minimizing an error function allows for the pose estimation directly from the matches between image point projections and their world coordinates. In this case the following error function can be used, with a minimum of 3 matched points coordinates,

$$F(X,\Theta) = \sum_{n=1}^{N} [d^2\left(x_n, \Psi(K,\Theta)\cdot x'_n\right) + d^2\left(x'_n, \Psi^{-1}(K,\Theta)\cdot x_n\right)] \quad (3)$$

where the pose parameters are represented by the 6-vector $\Theta = \begin{bmatrix} \alpha & \beta & \gamma & {}^{W}_{C}\mathbf{t}_x & {}^{W}_{C}\mathbf{t}_y & {}^{W}_{C}\mathbf{t}_z \end{bmatrix}$. If only the principal point and skew is known, then the same modification as above can be applied to the error function and the parameter vector, yielding

$$F(X,\Theta) = \sum_{n=1}^{N} [d^2\left(x_n, \Psi(u_0, v_0, fk_\theta, \Theta)\cdot x'_n\right) + d^2\left(x'_n, \Psi^{-1}(u_0, v_0, fk_\theta, \Theta)\cdot x_n\right)]$$

where
$\Theta = \begin{bmatrix} \alpha & \beta & \gamma & {}^{W}_{C}\mathbf{t}_x & {}^{W}_{C}\mathbf{t}_y & {}^{W}_{C}\mathbf{t}_z & fk_u & fk_v \end{bmatrix}$.

3.2.3. Self-Calibration

For the iterative estimation of the intrinsic parameters directly from the coordinates of matched points, an error function can be constructed using the particular structure of the homography between pure rotated views. The homography between two images captured by two cameras whose frames relate by the rotation matrix $R_{i+1,i}$ is $T_{i,i+1} = K.R_{i+1,i}K^{-1}$. By taking into account the distance errors on the coordinates of points matched over pairs of consecutive rotation images the following error function can be used,

$$F(X,\Theta) = \sum_{i=1}^{M-1}\sum_{n=1}^{N_i} [d^2\left(x^i_n, K.R_{i+1,i}K^{-1}\cdot x^{i+1}_n\right) + d^2\left(x^{i+1}_n, K.R_{i,i+1}K^{-1}\cdot x^i_n\right)]$$

where M is the number of images, N_i is the number of matched points between images i and $i+1$, and K and $R_{i,j}$ are functions of Θ. Here both the intrinsic parameters and the $M-1$ rotation matrices are simultaneously estimated. Under the assumption of constant intrinsics, a parameter vector allowing unconstrained minimization will be

$$\Theta = \begin{bmatrix} fk_u & fk_v & fk_\theta & u_0 & v_0 & \alpha_1 & \cdots & \gamma_{M-1} \end{bmatrix}$$

4. Propagating Uncertainty

We will now address the problem of estimating how the uncertainty is propagated over the various steps involved in the pose estimation. Although not considered during the mosaic creation

process, one of the sources of error in the estimated pose is due to the limited resolution of the feature extraction and matching, during the image-to-mosaic registration. In this work, we have considered that the perturbations affecting the coordinates of the matched points projections are the result of independent additive normal distributed noise with the same properties on every point projection.

Several authors have addressed the problem of estimating the error propagation in mosaicing applications. Recently, Kanatani[11] has presented a theoretically optimal algorithm for the computation of the homography between two images, using the framework of statistical optimization. Criminisi *et al.*[12] have shown how an homography estimate and the covariance of its elements can be used for measuring distances in the scene and estimate the associated uncertainties.

In our work, the methods used for the uncertainty propagation follow closely the one described by Haralick in [13]. In this paper, the author presents a general method for propagating the covariance matrix through any kind of linear or non-linear calculation, provided that a scalar function $F(X, \Theta)$ is defined which is minimized by the noisy calculation estimate $\widehat{\Theta}$ and noisy data $\widehat{X}$, and that the calculation can be well approximated by a first order Taylor series expansion for the level of noise involved.

An estimator for the covariance $\Sigma_{\Delta\Theta}$ of the noise in $\widehat{\Theta} = \Theta + \Delta\Theta$, is given by

$$\Sigma_{\Delta\Theta} = \left[\frac{\partial^2 F}{\partial \Theta^2}\left(\widehat{X}, \widehat{\Theta}\right)\right]^{-1} \left[\frac{\partial^2 F}{\partial X \partial \Theta}\left(\widehat{X}, \widehat{\Theta}\right)\right]^{T} \cdot \cdot \Sigma_{\Delta X} \cdot \frac{\partial^2 F}{\partial X \partial \Theta}\left(\widehat{X}, \widehat{\Theta}\right) \left[\frac{\partial^2 F}{\partial \Theta^2}\left(\widehat{X}, \widehat{\Theta}\right)\right]^{-T} \tag{4}$$

4.1. Experimental Validation

In order to test whether the error propagation for the functions defined above could be satisfactorily approximated by a first order series expansion, an experimental validation was carried out. The results regarding the case of pose estimation from matched image-to-world coordinates will now be presented.

A reference pose was chosen, from which two lists of noise-free coordinates of matches between the image and the world were obtained. A typical experimental level of 0.5 pixels for the standard deviation of the additive Gaussian noise was considered for creating instances of noisy image coordinates. It is here assumed that the noise involved in real imagery is caused by the limited resolution of the matching procedure and from slight non-planarities in the scene. Therefore, the considered level of noise was obtained by measuring the residuals resulting from the estimation of an homography between two real underwater images during the creation of the mosaic shown above.

The noise level affecting the world coordinates was set by measuring the statistics of the projections of the noisy image points, when back-projected onto the world, using the reference pose.

For each noisy instance of the matched coordinates, the corresponding pose was estimated using the minimizing criteria (3). The statistics of 500 pose instances where then compared with the predicted values, given by the estimator of Equation (4). The predicted covariance was computed around the mean value of the pose estimates. The Gaussian behavior of the pose estimate is depicted on Figure 2, with a superimposed normal density fit and predicted values. It can be seen that, for this level of noise the prediction is accurate.

An additional test was conducted, with different levels of noise, aimed at gaining insight on the limits of the approximation validity. In order to compare the real and the predicted covariance matrices, a criteria was devised based on the normalization of the real covariance matrix. By using Principal Components Analysis, one can find the linear transformation on the parameter space that maps the real covariance matrix onto the identity, provided the uncertainty spans all the parameter space. By applying the same transformation, both on the real and on the predicted covariance matrices, the criteria returns the Frobenius norm of their difference. The results for noise levels ranging from 0.25 to 20 pixel standard deviation and 500 instances each, are depicted on Figure 3. Based on this plot, one can conclude that the covariance prediction for the pose estimation, using matched coordinates, is accurate up to noise lev-

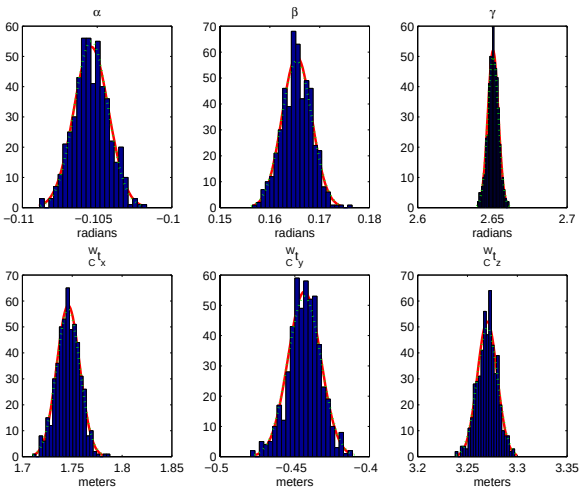


Figure 2. Testing results for the validity of the first order approximation in predicting the covariance. The pose parameter histograms where created with 500 instances of noise contamined coordinates, with a noise level of 0.5 pixels (standard deviation). The superimposed lines account for a normal density fit (full line) and predicted distribution (dotted line).

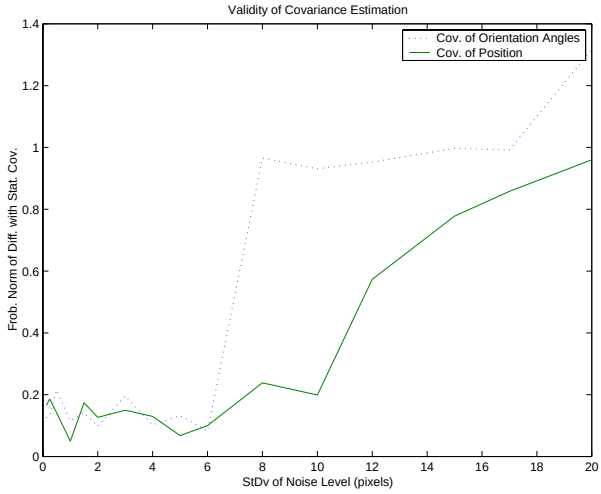


Figure 3. Evolution of the validity criterion for covariance estimation with increasing noise levels.

els of 6 pixel standard deviation, which is considerably higher than the experimentally measured noise of 0.5 pixel.

5. Pose Estimation Results

The performance of the pose estimation was experimentally evaluated by testing the camera trajectory reconstruction using ground-truth data. The test results presented in this section assume constant intrinsic parameters in time and differ both on the amount of intrinsic parameter information used, and on the nature of the observation data. Six methods have been implemented and tested. They are:

- Pose from $T_{i,World}$ with completely known intrinsic parameter matrix.
- Pose from $T_{i,World}$ with known principal point and skewing.
- Pose from $T_{i,World}$ with unknown intrinsic parameter matrix, but estimated using self-calibration from rotating scenes.
- Pose from image-to-World correspondences and completely known intrinsic parameter matrix.
- Pose from image-to-World correspondences and known principal point and skewing.
- Pose from image-to-World correspondences and unknown intrinsic parameter matrix, but estimated using self-calibration.

The third and the sixth methods build upon of the first and fourth, respectively. Under the former two, the self-calibration scheme described earlier is used to estimate the K matrix and its uncertainty. In practical applications, this implies the ability for an additional manoeuvre using a pan and tilt head mounted on the vehicle, or the possibility of the vehicle rotating maintaining the camera optical center approximately at the same position.

5.1. Experimental Setup

5.1.1. Original sequences

In order to evaluate the performance of the pose estimation algorithms, accurate ground-truth is required. For this reason we have used the mosaic of Figure 1 and synthesized new views according to a specified camera matrix and trajectory. These images are then used to retrieve the camera and position parameters. The mosaic was set to cover an area of 6 by 14.5 meters. The sequence comprises 40 images of 320×240 pixels taken by a camera on a moving vehicle combining 3D motion and rotation. The camera is pointing downwards with a tilt angle of approximately 150 degree with respect to the horizontal. The used intrinsic parameters matrix K accounts for a skewless camera with the following intrinsics,

$$K = \begin{bmatrix} 480 & 0 & 160 \\ 0 & 480 & 120 \\ 0 & 0 & 1 \end{bmatrix}$$

In order to simulate the vehicle drift induced by water currents, perturbations have been added to the nominal forward motion of 0.23 meters/frame and to a nominal height above sea floor of 3 meters. The perturbations account for periodic drifts of around 0.4 meters in position and 15 degrees in orientation. For each frame, the combined movement of the camera is depicted on Figure 4, where the camera is represented with its optical axis.

For the self-calibration method, an additional set of 20 images was produced, in which the camera undergoes pure rotation. The optical centre remained fixed at 4 meters above the sea bottom, while the camera faced down, and rotated around the 3 axes (pan, tilt and yaw). For each axis, the angle range is ± 5 degrees. The intrinsic parameters matrix K used for creating this sequence was the same as the one used for the other sequence.

The estimation of correspondences between adjacent images constitutes the starting point for the self-calibration procedure. The iterative self-calibration method described earlier was used for estimating both the intrinsic parameters and their associated uncertainty. Here a skewless, 4-parameter camera model was adopted during the minimization process. The recovered intrinsic parameter matrix which is used later on, is

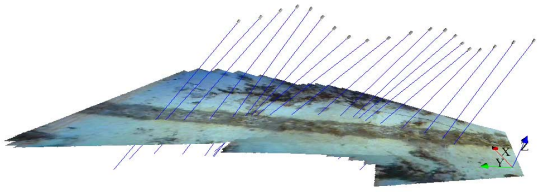


Figure 4. 3–D view of the camera positions and corresponding optical axes used for generating the sequence with available ground–truth. The origin of the 3–D world referential is represented by the system of three axes on the lower right, where each axis is drawn at 1 meter length.

$$K_{rec} = \begin{bmatrix} 479.0 & 0 & 158.7 \\ 0 & 487.7 & 125.7 \\ 0 & 0 & 1 \end{bmatrix}$$

5.1.2. Algorithm for new registration on the mosaic

Next, we will deal with the problem of finding point correspondences relating a sequence of newly acquired images to a previously constructed mosaic. In this, we explicitly take advantage of the timely order nature of the image sequence, to reduce the computational burden of finding correspondences on the mosaic. By assuming sufficiently large image overlap over adjacent image frames, the feature matching search can be restricted to a neighboring area of the location predicted by the last image-to-mosaic homography. Also image warping of the feature is performed before correlation.

The implemented algorithm requires an estimate of the first image-to-mosaic homography, which needs not to be very accurate. For each image, the algorithm tries to find a reliable image-to-mosaic correspondences set. Reliability is insured by the specification of a minimum acceptable number of correct matches. If it fails to find that number, the algorithm uses the homography with the previous image to compute an approximation of the current image-to-mosaic homogra-

phy, in order to narrow the search area for the correspondences. The advantage of registering each frame directly on the mosaic (as opposed to computing by sequentially cascading the homographies $T_{i-1,i}$, between previous images), is due to the fact that small estimation errors on $T_{i-1,i}$ are not accumulated.

Once the image-to-mosaic correspondences have been found and the mosaic is referenced to a world frame, the camera pose can be estimated with respect to the selected world frame.

5.2. Pose Estimation Results

The synthetic images from the sequence containing camera translation were registered directly on the mosaic, using the algorithm described above. This algorithm was run with a specified acceptable minimum of eight matched pairs per homography. In each frame it was able to find between 16 and 39 pairs. For the set of 40 images, 2 homographies were computed with matched pairs from a second attempt, while the other 38 were computed at the first attempt.

For all experiments, the uncertainty in the image point correspondences was modeled as additive Gaussian noise, independent for each coordinate. A value of 0.5 pixel standard deviation was used as the input for the covariance prediction of the elements of the homographies and pose. This value was chosen as a conservative approximation to the real standard deviation that was measured from the residuals of an homography computation between two real underwater images.

For each of the six methods, an initial estimate was found using the non-iterative solutions provided by the corresponding methods of Section 3.1. Next, a quasi-Newton nonlinear optimization algorithm[14] was used for minimizing the corresponding cost function as detailed in Section 3.2.

The pose covariance prediction in all the methods takes into account the uncertainty on the intrinsic parameters, being these provided beforehand or estimated using self-calibration. However, for the experiments reported here, only the methods using self-calibration take into account the predicted uncertainty for the K matrix. The experiments for the other methods, assuming known K and known principal point, use error-free intrinsic parameters.

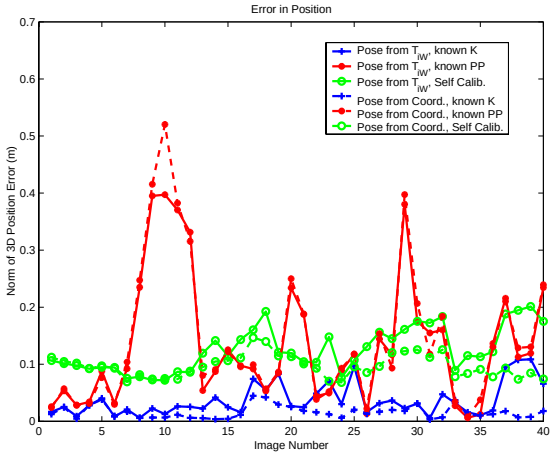


Figure 5. Position error for all the pose recovery methods using mosaic registration.

Statistics for the reconstruction errors are presented in Table 1. The position errors were measured by taking the Euclidean distance between the ground-truth position and the estimated position. As for the orientation, the error was measured by computing the angle between the true and estimated camera frame orientations. For each image and method, the position errors are plotted on the left side of Figure 5. On the right side, the predicted uncertainty is represented by the volume of the ellipsoid accounting for 50% of the uncertainty.

In these results, the lowest position and orientation errors correspond to the trajectory recovery methods for known K matrix. This is not surprising, as these methods use the most prior information. The second best class of methods are for the pose estimation using self-calibration for which the average distance of the position error is four times the one of the previous methods. The methods yielding poorer results in terms of average error are the ones using just the principal point information.

Within each class of intrinsic information used, the methods presenting the best results are the

Method	Position Errors (meters)		Angular Errors (degrees)	
	Avg. Norm	Avg. Unc. ($\times 10^{-3}$)	Avg. Norm	Avg. Unc. ($\times 10^{-3}$)
Pose from T_{iW}, known K	0.038	0.185	0.350	0.001
Pose from Coord., known K	0.016	0.051	0.252	0.001
Pose from T_{iW}, known PP	0.135	1.558	1.151	0.003
Pose from Coord., known PP	0.142	0.315	1.209	0.004
Pose from T_{iW}, Self–Calib.	0.126	0.482	0.814	0.002
Pose from Coord., Self–Calib.	0.097	0.116	0.982	0.003

Table 1
Trajectory recovery results for the methods using known K matrix, known principal point, and self–calibration. The average norm refers to the mean value of the error distances, while the average uncertainty refers to the mean value of the 50% uncertainty volume.

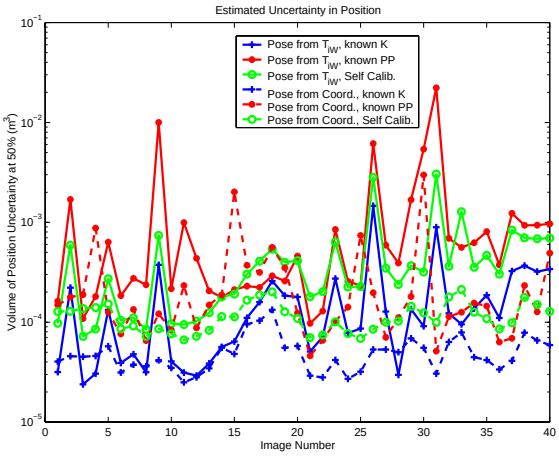


Figure 6. Estimated uncertainty for the position for all the pose recovery methods using mosaic registration.

ones estimating the pose from point correspondences instead of using the intermediate image-to-World homography, with the exception of the methods using known principal point for which the average error norm is approximately the same. A justification for this condition lies on the fact that a larger set of observation data is used in pose from coordinates when compared with the pose from the $T_{i,World}$ homography.

For all six methods, the corresponding 3-D views of the recovered trajectories are depicted in Figure 7. Again, it can be seen that the methods using the $T_{i,World}$ homography produce larger uncertainty volumes. Also, for certain positions and orientations, the uncertainty volumes are much larger than the average. Although not experimentally verified, this condition is likely to be the result of pose configurations where the used parameterization for $T_{i,World}$ amplifies the noise in certain directions of the space spanned by its components.

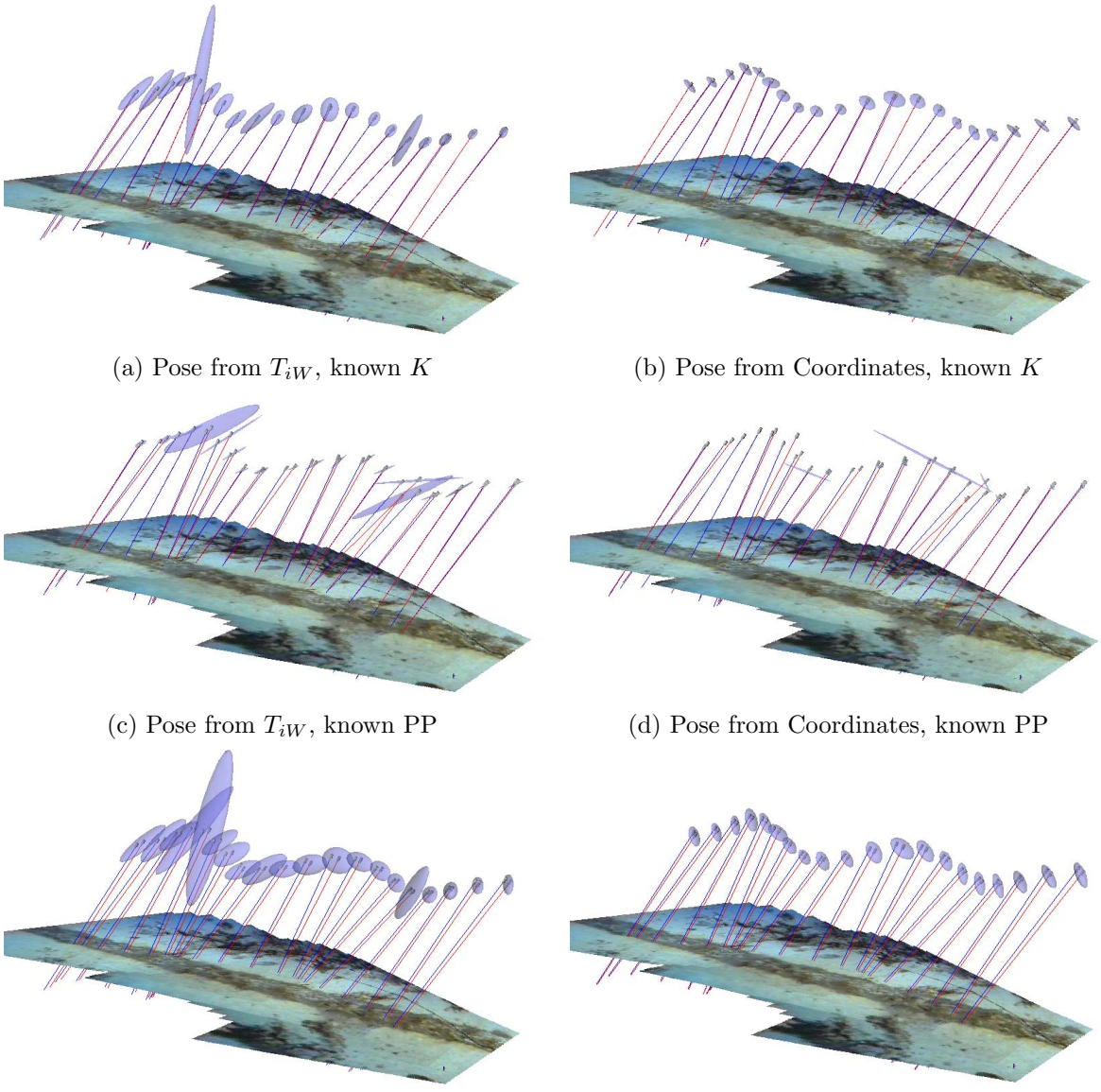
5.3. Pose from inter-image homographies

An additional experiment was conducted in order to compare the following image registration schemes:

- Image-to-mosaic homographies computed by direct mosaic registration

- Image-to-mosaic homographies computed by cascading inter-images homographies

The first scheme refers to the first method of Section 5.2, which makes use of the mosaic regis-



(a) Pose from T_{iW}, known K

(b) Pose from Coordinates, known K

(c) Pose from T_{iW}, known PP

(d) Pose from Coordinates, known PP

(e) Pose from T_{iW}, Self Calib.

(f) Pose from Coordinates, Self Calib.

Figure 7. VRML views of the estimated trajectory positions and uncertainty ellipsoids for the pose recovery experiments. Only one out of every two recovered camera positions, is plotted. The original camera axes are drawn in a darker colour (blue), while the recovered camera axes are drawn in a lighter colour (red). The size of the ellipsoids was set for a 50% probability. For better visual perception, the ellipsoids for methods (a), (b), (e) and (f) have been enlarged by 5 times, while for (c) and (d) they retain the original size.

tration algorithm described in 5.1.2. In the second, the true camera position and orientation is used for computing the first image-to-mosaic homography $T_{M,1}$. The subsequent homographies are calculated by,

$$T_{M,i} = T_{M,1} \cdot \prod_{k=2}^{i} T_{k-1,k} \qquad i > 1$$

where $T_{k-1,k}$ are the inter-images homographies and the matrix product is computed by right-multiplying for each increment of the index k. The set of $T_{k-1,k}$ was estimated from the same sequence of images, and the number of used matched points varied from 10 to 76 pairs, with an average of 60.

Figure 8 and Figure 9 present, respectively, the plot of the positions errors for each frame, and a 3-D reconstruction of the two trajectories. It can be seen that the second scheme produces much less accurate results, due to the fact that small errors, inherent to the inter-image homography estimation, are accumulated. This phenomena is in many ways comparable to the positioning errors arising from the use of dead-reckoning during navigation.

6. Conclusions

We have presented an approach for the use of underwater video mosaics as visual reference maps for vehicle localization. Key issues for the mosaicing process are the robust selection of correspondences and the use of geometric models capable of registering any view of a planar scene.

Methods for pose estimation were presented, which allow the estimation of the 3D position and orientation of a vehicle from a new view of a previously created mosaic. For each method, the associated uncertainty in the pose parameters, as a function of the uncertainty in the observations, was implemented using a first order approximation. For the levels of noise involved in the experimental part of this work, the approximation was validated by the good fit between the predicted and measured statistics. Apart from providing criteria for comparing different methods, the importance of the covariance prediction

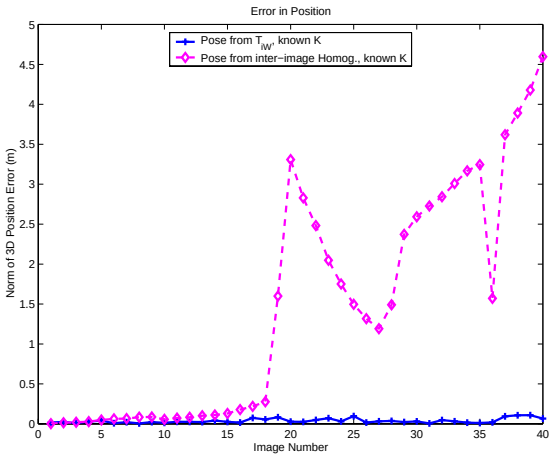


Figure 8. Position error for the pose recovery methods using direct mosaic registration, and inter–image homography cascading.

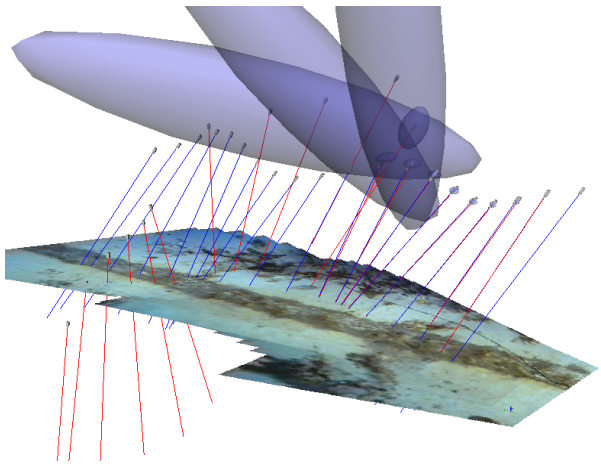


Figure 9. Estimated trajectory positions and uncertainty ellipsoids for pose recovery using inter–image homographies. Only one, out of every two recovered camera positions, is plotted. The ellipsoids are set for a 50% probability, but due to their rapid growth, only the first half are drawn.

is also apparent from the fact that it can be used for the detection of degenerate configurations and it enables the monitoring of position uncertainty during navigation.

The different pose estimation methods were evaluated using an image sequence with ground-truth. The performance was compared both in terms of pose error and in terms of predicted estimate covariance.

An emphasis was put on using several degrees of available information on the camera intrinsic parameters, including self-calibration. The possibility of calibrating a camera *on-line* can be of practical importance for a number of visually guided tasks, specially if the camera parameters are subject to change slowly in time. This paper illustrated how relevant information for the pose estimation process can be obtained by the analysis of rotation images. These images are easier to acquire than having to resort to calibration grids.

By automatically creating visual representations of the sea floor and using them for navigation, the methods described in this paper provide an important step towards the autonomous operation of submersibles.

Acknowledgments

The work described in this paper has been supported by the Portuguese Foundation for Science and Technology PRAXIS XXI BD/13772/97, and by NARVAL Esprit-LTR Proj. 30185.

REFERENCES

1. J. Santos-Victor, J. Sentieiro, The role of vision for underwater vehicles, in: Proc. of the 1994 Symposium on Autonomous Underwater Vehicle Technology, Cambridge, MA, USA, 1994, pp. 28–35.
2. N. Gracias, J. Santos-Victor, Automatic mosaic creation of the ocean floor, in: Proc. of the IEEE OCEANS'98, Nice, France, 1998.
3. X. Xu, S. Negahdaripour, Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images, in: Proc. of the Oceans '97 Conference, Vol. 2, Halifax, Canada, 1997, pp. 1412–17.
4. J. Zheng, S. Tsuji, Panoramic representation for route recognition by a mobile robot, International Journal of Computer Vision 9 (1) (1992) 55–76.
5. J. Batista, H. Araújo, A. Almeida, Iterative multi–step explicit camera calibration, in: Proc. of the 6th International Conference on Computer Vision, Bombay, India, 1998.
6. S. Ganapathy, Decomposition of transformation matrices for robot vision, in: Proc. 1st IEEE Conf. Robotics, IEEE, 1984, pp. 130–139.
7. N. Gracias, J. Santos-Victor, Underwater video mosaics as visual navigation maps, Computer Vision and Image Understanding .
8. N. Gracias, Application of robust estimation to computer vision: Video mosaics and 3–D reconstruction, Master's thesis, http://www.isr.ist.utl.pt/labs/vislab/thesis, Lisbon, Portugal (April 1998).
9. R. Hartley, Self–calibration from stationary cameras, International Journal of Computer Vision 22 (1) (1997) 5–23.
10. R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, 1985.
11. K. Kanatani, N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, in: Proceedings of the Seventh International Conference on Computer Vision, Vol. 1, IEEE, 1999, pp. 73–78.
12. A. Criminisi, I. Reid, A. Zisserman, A plane measuring device, Image and Vision Computing 17 (8) (1999) 625–634.
13. R. Haralick, Propagating covariance in computer vision, in: Proc. of the Workshop on Performance Characteristics of Vision Algorithms, Cambridge, UK, 1996.
14. W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 1988.