Camera Self-Calibration in Underwater Environment

Nathalie Pessel, Jan Opderbecke, Marie-José Aldon

Camera Self-Calibration
in underwater environment

Nathalie PESSEL
IFREMER[1]
Zone portuaire de Brégaillon
83507 La Seyne-sur-mer
France
Nathalie.Pessel@ifremer.fr

Jan OPDERBECKE
IFREMER[1]
Zone Portuaire de Brégaillon
83507 La Seyne-sur-mer
France
Jan.Opderbecke@ifremer.fr

Marie-José ALDON
LIRMM[2]
161, rue ADA
34392, Montpellier
France
Marie.Jose.Aldon@lirmm.fr

ABSTRACT

This paper presents a self-calibration technique for a camera mounted on an underwater vehicle designed to perform the 3D reconstruction of underwater scenes. Our aim is to identify the intrinsic parameters of the camera with methods that are adapted to the operational constraints on Ifremer's underwater vehicles.

The optical system is composed by a single vertical camera located below the underwater vehicle and looking downwards. The motion of the vehicle can be measured through navigation sensors and the observed 3D scene is always unknown. The use of a moving camera is not an obstacle for the application of stereoscopic methods. Nevertheless, the camera motion enables the use of robust algorithms for points matching, but impoverishes perspective effects between several images. Therefore, we are interested in the analysis of the conditions in which the procedure of self-calibration is valid and reliable, i.e.: the 3D characteristics of the scene and the camera motion. This paper presents the steps necessary for the camera self-calibration in an underwater environment: the extraction and the tracking of features in several successive images, the fundamental matrix estimation and the intrinsic parameters identification. Several tests and results are presented.

Keywords

Self-calibration, intrinsic parameters, fundamental matrix, matching, tracking, RANSAC

1. INTRODUCTION

Underwater vision is an essential element in subsea operations which rely nowadays mostly on an unmanned underwater vehicles. In order to quantitavely exploit underwater images in the varying optical conditions of the subsea environment, tools for rapid automated camera calibration without external means are required.

We develop in this paper a method for calibrating a camera that is mounted in a fixed position on an underwater vehicle. The method aims to identify the intrinsic parameters of a camera, given an image sequence corresponding to a specific vehicle trajectory for a significant natural scene. The extrinsic parameters of the camera are measured or estimated by the navigation system (gyro compass, Doppler log, acoustic altimeter, pressure sensor, …).

The most common methods for camera calibration use a specific object (e.g. a planar calibration grid), in order to identify known dimensions and to simplify feature matching between the two images. Furthermore, most applications consider the use of a stereo pair of cameras.

The self-calibration method that we propose uses images from a dense sequence with small displacements between successive frames. We can determine point matches between two images A and B with significant perspective change by tracking feature points through the image sequence from A to B. As our calibration algorithm does not imply observation of a scene of known dimensions, there is no necessity to deploy a specific device.

As the images used for calibration are provided by one moving camera, the movement underlying the sequence is of prime importance. Experiments have allowed to determine the most significant trajectory, which takes into account the constraints and controllability of the vehicle's movements.

We design in the following paragraphs all algorithmic steps required for the self-calibration of a camera. At first, we present the extraction and the tracking of points in the image sequence. Section 3 presents the epipolar geometry and more precisely the estimation of the fundamental matrix. The method for computing the intrinsic parameters is explained in section 4. Section 5 illustrates experimental results obtained by using the implemented algorithms.

2. EXTRACTION AND MATCHING OF POINTS

Most of classical stereoscopic methods distinguish between extraction and matching of features. In this respect, Deriche [Der90] uses contour points to identify the features, Harris [Har88] considers interest points and Sistiaga [Sis00] local invariants. The features to be matched are extracted from both images and matched using correlation techniques or a measure of distance between differential attribute vectors.

In our application where image sequences feature little displacement between two successive images, a tracking algorithm based on the principles developed by Kanade-Lucas-Tomasi (KLT) [Tom91] is found to be more appropriate. As opposed to the methods mentioned above, the KLT algorithm has the particularity to extract characteristics only in a first image and to track them through a short image sequence.

The algorithm's properties will be outlined while discussing their role in our application.

2.1 Extraction of points

A selected primitive is a textured patch with a large variation of intensity in the two directions x and y.

Given the intensity function $I(x, y)$, the matrix of variation of local intensity Z is:

$$Z = \int_W \mathbf{g}\mathbf{g}^T dW = \int_W \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} dW \qquad (1)$$

where $\mathbf{g}$ is the local gradient defined as: $\mathbf{g} = \left(\dfrac{\partial I}{\partial x}, \dfrac{\partial I}{\partial y} \right)$

and I_x and I_y are the first derivatives of the function I along directions x and y.

A patch defined by a squared window W (for example 15 pixels) is accepted as an interesting primitive if the two eigenvalues (λ_1, λ_2) of Z are superior to a threshold value λ:

$$\min(\lambda_1, \lambda_2) > \lambda \qquad (2)$$

The matrix Z can be computed in function of x and y by displacing the window W by a number of pixels smaller than the window size.

For each displacement of the window, eigenvalues of the matrix Z are computed. Thresholding according to equation (2) gives a set of window locations that represent feature points.

2.2 Feature tracking

Feature tracking is based on the assumed displacement d between the two images. Features can be tracked using a measure of similarity, if the residual displacement between two successive images I and J is smaller than the observation window. As two images are extracted from a continuous sequence, the displacement is assumed to be small and can be approximated by a translation. The image can then be defined as a function of three variables (x, y, t):

$$I(x, y, t) = I(x + \xi, \ y + \eta, \ t + \tau) \qquad (3)$$

where the displacement between times t and $t + \tau$ at point $\mathbf{x} = (x, y)$ is $d = (\xi, \eta)$.

We can write:

$$I(\mathbf{x} + d) = I(x + \xi, \ y + \eta, \ t + \tau)$$

A second image can then be written as:

$$J(\mathbf{x}) = I(\mathbf{x} + d) + n(\mathbf{x}) \qquad (4)$$

where n is a function representing the image noise.

The problem is to determine d by minimizing the dissimilarity between the two windows W in I and J.

$$\varepsilon = \int_W [I(\mathbf{x}+d) - J(\mathbf{x})]^2 \, w(\mathbf{x}) \, d\mathbf{x} \qquad (5)$$

When the displacement vector d is small, the intensity function can be approximated by a Taylor series truncated to the linear term:

$$I(\mathbf{x}+d) = I(\mathbf{x}) + \mathbf{g} \cdot d \qquad (6)$$

where $\mathbf{g}$ corresponds to the image gradient.

According to [Shi94], we have:

$$Z\mathbf{d} = e \qquad (7)$$

where :

- Z can be expressed by the truncated Taylor series:

$$Z = \int_W \mathbf{g}(\mathbf{x}) \, \mathbf{g}^T(\mathbf{x}) \, w(\mathbf{x}) \, d\mathbf{x} \qquad (8)$$

- e is a column vector representing the difference between the two images:

$$e = \int_W [I(\mathbf{x}) - J(\mathbf{x})] \mathbf{g}(\mathbf{x}) \, w(\mathbf{x}) \, d\mathbf{x} \qquad (9)$$

The displacement d is therefore computed by solving the system (7).

3. EPIPOLAR GEOMETRY BETWEEN TWO SUCCESSIVE IMAGES

The projective geometry corresponding to two images of a same scene taken in two different view points is called epipolar geometry. It can be represented algebraically by the fundamental matrix F.

3.1 Fundamental matrix

By definition, the fundamental matrix is associated to the transformation which links a point of the left image to a line containing its corresponding point in the right image. The F matrix is function of intrinsic and extrinsic parameters of the camera and arises in the epipolar constraint in the following way:

$$q_i'^T F q_i = 0 \quad \forall i \in [1, n] \qquad (10)$$

Several methods exist to identify the fundamental matrix parameters. One of these methods [Fau92] consists in minimizing at the same time in the two images, the sum of distance squares from a point to the

epipolar line which is supposed to pass by this point. A further approach is based on the minimization of (10) by weighing the gradient of its variance. On the contrary of these two methods which are non-linear, Hartley [Har95] proposes to estimate the fundamental matrix using a linear method.

3.2 Fundamental matrix estimation

In the approach presented here, the estimation of the fundamental matrix is realized by using Hartley's normalized 8-point algorithm [Har95].

Using this method the problem is solved by first normalizing the matched point coordinates. During this stage, each image reference frame is first translated to the centroïd of the set of all points. Then, an isotropic scaling of points allows us to reduce the distance at the origin so that the average value of the distances is equal to $\sqrt{2}$. The normalized 8-points algorithm allows us to compute the fundamental matrix from a set of at least eight matched point $q_i' \leftrightarrow q_i$. In particular, writing $q = (u, v, 1)^T$ and $q' = (u', v', 1)^T$ each point match gives rise to one linear equation in the unknown entries of F :

$$
\begin{aligned}
&u\,u'\,f_{11} + u\,v'\,f_{21} + u\,f_{31} + v\,u'\,f_{12} + \\
&v\,v'\,f_{22} + v\,f_{32} + u'\,f_{13} + v'\,f_{23} + f_{33} = 0
\end{aligned}
\qquad (11)
$$

This linear equation can be written in the form:

$$U^T f = 0 \qquad (12)$$

where: $\quad U = [u\,u', u\,v', u, v\,u', v\,v', v, u', v', 1]$
$$f = [f_{11}, f_{21}, f_{31}, f_{12}, f_{22}, f_{32}, f_{13}, f_{23}, f_{33}]^T$$

If we have n correspondences, we obtain the following matrix equation:

$$
Af = \begin{bmatrix}
u_1 u_1' & u_1 v_1' & u_1 & v_1 u_1' & v_1 v_1' & v_1 & u_1' & v_1' & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
u_n u_n' & u_n v_n' & u_n & v_n u_n' & v_n v_n' & v_n & u_n' & v_n' & 1
\end{bmatrix} f = 0
$$
$$(13)$$

The fundamental matrix F is defined only up to an unknown scale factor. For this reason and to avoid a trivial solution of f, a constraint was added: $\|f\| = 1$.

Eight point correspondences are therefore necessary to linearly compute F. If the data are noisy and inaccurate, as it is often the case in practice, the

estimation of F is obtained by using linear least squares. Thus, we seek to estimate the f vector that minimizes $\|Af\|$. The solution f of this problem is the eigenvector corresponding to the smallest eigenvalue of AA^T.

An important constraint is associated with the fundamental matrix. It has to be of rank 2. Therefore the fundamental matrix F is replaced by the matrix F' that minimises the Frobenius norm $\|F - F'\|$ subject to the condition: $\det(F') = 0$. The F' matrix is the closest singular matrix to F under Frobenius norm.

We have shown that the estimation of the fundamental matrix F can be obtained through point matching. The presence of false matches perturbs the estimation of F, therefore the elimination of bad point matches is essential. In order to do this, we suggest the use of the RANSAC (RANdom Sample Concensus) algorithm as developed by Fischler and Bolles [Fis81].

3.3 RANSAC algorithm

The validation of matches comes down to segment data in good and false matches [Tor93].
The working principle of this algorithm is to define a model and to determine it with N random samples of n points. From these N determinations of the model, a classification of matches (as valid or not) is possible by using a criterium of points validity.
It is not necessary to try every possible sample of points. The number of samples N is chosen sufficiently high to ensure with a probability p, that at least one of the random samples of n points consists solely of valid matches. N is given by:

$$N = \log(1-p) \big/ \log(1-(1-\varepsilon)^n) \qquad (14)$$

with $p = 0.99$ and ε the probability that a selected point is an "invalid point".
For the camera intrinsic parameters estimation, the RANSAC algorithm is coupled with the computation of the fundamental matrix between two successive images. The model, which represents the mathematical support allowing to verify the point matches is also expressed by the computation of the fundamental matrix.
The points validity characterization criterium is a threshold which defines the accuracy of the point matching. In order to verify the validity of a matching, this threshold is compared to the error obtained when reprojecting the matched point by a triangulation

algorithm [Har97]. The threshold varies in function of the wanted precision. In our application, the threshold is defined equal to 1 or 2 pixels. For that reason, a matching is considered valid if the two corresponding points in the two left and right images are removed from real pixels of a distance inferior or equal to the threshold.

4. INTRINSIC PARAMETERS ESTIMATION

The camera's intrinsic parameters are calculated by using the Mendonça and Cipolla method [Men99] applied to a set of five images taken at given intervals from a dense sequence.

Mendonça and Cipolla proposed a method based on the the essential matrix properties. The three singular values of the essential matrix have to satisfy two conditions: one of them must be zero and the two others must be equal.

For that, the authors proposed the use of a cost function which takes the intrinsic parameters as arguments and the fundamental matrix as parameters. This function minimizes a positive value proportional to the difference between the two non-zero singular values of the essential matrix.

The essential matrix is given by:

$$E = [t]_x R \qquad (15)$$

with: $[t]_x$: antisymmetric matrix associated to the translation vector t,
R : rotation matrix.

Moreover, we know that the expression of the fundamental matrix is (with K the matrix of intrinsic parameters):

$$F = K^{-T} [t]_x R K^{-1} \qquad (16)$$
$$F = K^{-T} E K^{-1} \qquad (17)$$

then $\quad K^T F K = E \qquad (18)$

Let F_{ij} be the fundamental matrix associated to consecutive images i and j, and $^1\sigma_{ij} > {}^2\sigma_{ij}$ be the non-zero singular values of E_{ij} obtained by making a singular value decomposition (SVD) of E_{ij}. The cost function is:

$$C(K) = \sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij} \frac{{}^{1}\sigma_{ij} - {}^{2}\sigma_{ij}}{{}^{2}\sigma_{ij}} \qquad (19)$$

with: - $K = f(\alpha_u, \alpha_v, u_0, v_0)$ where α_u, α_v, u_0, v_0 correspond respectively to products of the scale factors according to the axis u and v by the focal and to the coordinates of the intersection of the optical axis with the image plane,

   - w_{ij} is the degree of confidence of the fundamental matrix F_{ij} estimation.

5. EXPERIMENTAL RESULTS

Results presented below are organized into three parts. First we present results of the tracking method applied to two sequences of underwater images. Secondly we discuss robust parameter estimation using the RANSAC algorithm. At last, we present quantitative results for different kinds of movements, which allow to determine a significant trajectory.

5.1 Tracking of points

Given a sequence of images, characteristic points are extracted in the first image and tracked through the images of the sequence. Below we show this process (figure 1 & 2) with the tracked points drawn in the first and last frame of a five image sequence.

The first sequence corresponds to an almost linear motion with both vertical and horizontal components in the image plane. These images present two very distinct zones: the top part consists of a sandy ground and the lower part shows amphoras.
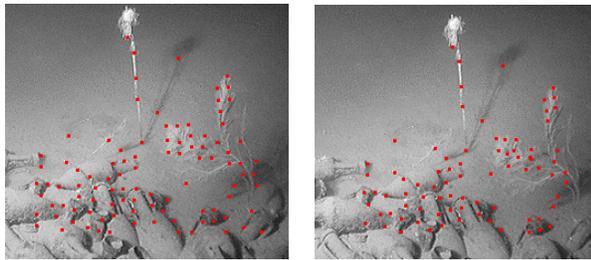
1a: image 1   1b: image 5

Figure 1: sequence "amphoras"

In figure 1a and 1b the correspondences between 82 extracted features shows the robustness of the method. Points are correctly tracked in the whole scene, despite the fact that the most relevant part of the scene is concentrated at the bottom of the image. Only 5 false matches are identified in the last image of the sequence.

The second sequence consists of images of a coral reef taken during a vertical translation in the image plane. These images show only a part of the images sequence.
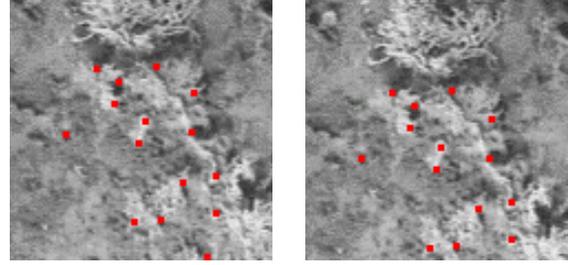
2a: image 1   2b: image 5

Figure 1: sequence "coral reef"

Despite the complexity of the image, 14 features are positively tracked through the sequence (Figure 2b). The points are well distributed in the image.

The experiments carried out on the two image sequences confirm the efficiency and the robustness of the KLT algorithm while processing underwater images.

5.2 Robust parameter estimation

This part of the study aims at analyzing the influence of the number of point matches on the accuracy of parameter estimation.

This study is a statistical comparative study of the camera self-calibration showing also the interest of using the RANSAC algorithm. We realized 100 runs with independent noise applied to the heather root sequence. The number of matched points varied from 10 to 90. The following figures present errors expressed as a percentage of parameter value in function of the number of point matches.

a: $\varepsilon_{\alpha u}$ b: $\varepsilon_{\alpha v}$

- - - without RANSAC
—— with RANSAC
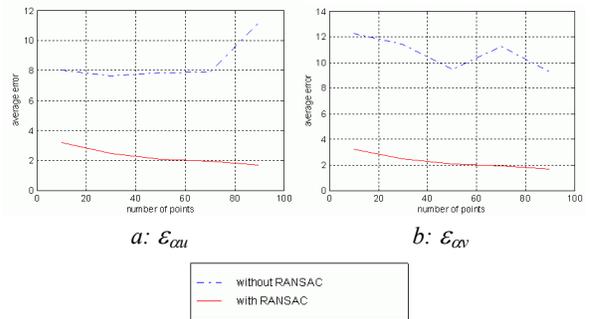
Figure 3: Errors in parameters α_u and α_v estimations in function of number of points used

a: ε_{u0}   *b: ε_{v0}*

- - - without RANSAC
—— with RANSAC
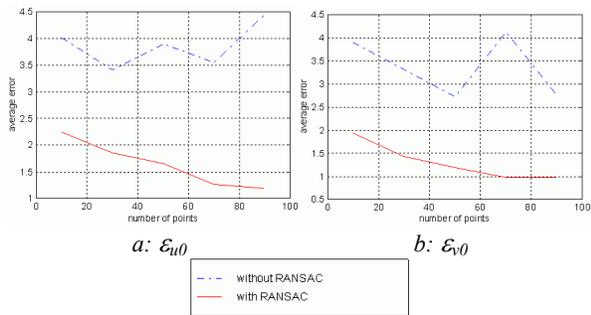
Figure 4: Errors in parameters u_0 and v_0 estimations in function of number of points used

On one hand, we observe that the integration of the RANSAC algorithm in the process of self-calibration enables to reduce errors significantly. As a compromise between processing time and estimation performance, we limit the number of point matches to 50.

For this example, the errors in intrinsic parameter estimation are: $\varepsilon_{\alpha u} = 2.05\%$, $\varepsilon_{\alpha v} = 2.05\%$, $\varepsilon_{u0} = 1.65\%$ and $\varepsilon_{v0} = 1.19\%$. These results are satisfying.

5.3 Camera movement vs. quantitative results

Several experiments were performed in order to evaluate the estimation of intrinsic parameters of the camera using the complete algorithm scheme presented in sections 2-4.

The experiments feature the use of several types of camera motion and several observed geometries, in conditions which are realistic for an underwater vehicle. In fact, all vehicle motions are controllable except the rotations around the horizontal axes (pitch and roll angles). However, these motions always present to some small amount are measurable. We began the experiments applying only the controllable motions (for example rotation around the optical axis). After that, small pitch and roll angles ($\pm 2°$) have been added.

The performed experiments can be grouped into two sets respectively relating to 2D scenes (plane) and 3D scenes. For each of them, we used real and simulated data.

The experiments relating to the real scenes were carried out using a digital camera.

The objects used during acquisitions of real sequences are presented in figure 5: (*a*) a 2D planar calibration grid, which has a regular geometric shape but no texture and (*b*) roots of heather, which represent a natural irregular but textured 3D shape.

a   *b*
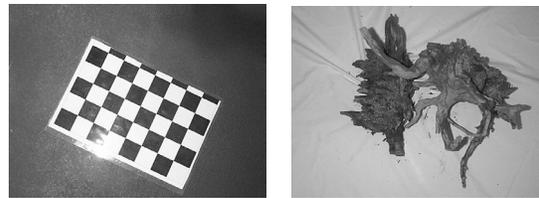
Figure 5: objets used in the experiments

The experiments related to the simulated scenes were realized by computing the images of 3D points belonging to a sphere (for the 3D simulations) and to a circle (for the 2D simulations). The camera's intrinsic parameters were fixed as follows: $\alpha_u = \alpha_v = 500$ and $u_0 = v_0 = 600$. The image size considered is 1200x1200 pixels. The simulations were carried out while introducing a noise with a standard deviation of 2 pixels in the coordinates of the calculated image points. The experiments were realized with sequences composed of five successive images of the same scene. Fifty points were tracked in each sequence.

The table below presents errors expressed as a percentage of the parameter value estimated in function of the object and of the camera motion.

Move-ments	2D scene		3D scene	
	Simulated points	Calibration grid	Simulated points	Roots of heather
θ_z	$\varepsilon_{\alpha u} =$ 66.6 % $\varepsilon_{\alpha v} =$ 66.6 %	$\varepsilon_{\alpha u} =$ 24.90 % $\varepsilon_{\alpha v} =$ 24.89 %	$\varepsilon_{\alpha u} =$ 66.6 % $\varepsilon_{\alpha v} =$ 66.6 %	$\varepsilon_{\alpha u} =$ 27.43 % $\varepsilon_{\alpha v} =$ 28.73 %
θ_z + t_z	$\varepsilon_{\alpha u} =$ 17.93 % $\varepsilon_{\alpha v} =$ 17.93 %	No result	$\varepsilon_{\alpha u} =$ 27.476% $\varepsilon_{\alpha v} =$ 27.51 %	No result
θ_z + (θ_x, θ_y)	$\varepsilon_{\alpha u} =$ 0.031 % $\varepsilon_{\alpha u} =$ 0.038 %	$\varepsilon_{\alpha u} =$ **0.26 %** $\varepsilon_{\alpha v} =$ **0.48 %**	$\varepsilon_{\alpha u} =$ 0.067 % $\varepsilon_{\alpha u} =$ 0.068 %	$\varepsilon_{\alpha u} =$ **2.06 %** $\varepsilon_{\alpha v} =$ **2.06 %**
t_v + (θ_x, θ_y)	No result		$\varepsilon_{\alpha u} =$ 34 % $\varepsilon_{\alpha v} =$ 0.961 %	$\varepsilon_{\alpha u} =$ 7.53 % $\varepsilon_{\alpha v} =$ **1.59 %**

Table 1: Errors in intrinsic parameter estimations
θ_z: rotation around the optical axis,
θ_x: roll angle and θ_y: pitch angle

The experiments, briefly resumed in table 1, led to the following conclusions:

- translation along y with varying pitch and roll angles allows us to estimate α_v which is the scale factor according to the horizontal axis of the image; we can deduce that an x translation with varying pitch and roll angles will allows us to find again the scale factor according to vertical axis of the image; coordinates of the optical center can not be estimated from translation;
- rotation around vertical axis accompanied or not by vertical translation shows poor estimation results;
- rotation around the vertical axis leads to positive results when introducing small but non-zero roll and pitch movement; this can be explained by the roll and pitch angles emphasizing the 3D geometry of the scene; note that roll and pitch motion is not controllable, but we can accentuate them by linear accelerations.

After these experiments, we can conclude that the best exciting trajectory for the intrinsic parameter estimation in our application, is a rotation around the optical axis with pitch and roll angles. The differences in results between the sequence of the planar calibration grid and that of the heather are due to the precision and to the quality of matches.

6. CONCLUSION

The camera self-calibration method described here presents some advantages compared to classical methods for our applicative conditions.

All algorithm steps have been validated with simulated and real data.

A first condition for successful self-calibration relies on representative point matches. We choose to track a considerable number of points and to use a robust identification method (RANSAC).

The most important issue is the camera movement. We have determined camera trajectories leading to good calibration results.

A future study will aim at the development of numerical criteria to indicate significant movement allowing self-calibration.

7. REFERENCES

[Der90] Deriche, R. and Faugeras, O., 2D Curve matching using high curvature points : Application to stereo vision, In Proceedings International Conference on Pattern Recognition, pp. 240-242, 1990.

[Fau92] Faugeras, O., Luong, Q-T. and Maybank, S., Camera Self-Calibration : Theory and Experiments, In Proceedings European Conference on Computer Vision, pp. 321-334, Santa-Marguerita, Italy 1992.

[Fis81] Fischler, M. A. and Bolles, R. C., Random Sample Concensus: a paradigm for model fitting with application to image analysis and automated cartography, Communication Association and Computing Machine, 24(6), pp. 381-395, 1981.

[Har88] Harris, C. and Stephens, M., A combined corner and edge detector, In Proceedings fourth Alvey Vision Conference, pp 174-151, Manchester, England, August 1988.

[Har95] Hartley, R., In Defence of the 8-points Algorithm, In Proceedings Fifth International Conference on Computer Vision, Cambridge, Mass., June 1995.

[Har97] Hartley, R. and Sturm, P., Triangulation, In Computer Vision and Image Understanding, 68(2), pp. 146-157, 1997.

[Men99] Mendonça, P. and Cipolla, R., A Simple Technique for Self-Calibration, In Conference on Computer Vision and Pattern Recognition, 1999.

[Shi94] Shi, J. and Tomasi, C., Good Features to Track, In IEEE Conference on Computer Vision and Pattern Recognition, Seattle, June 1994.

[Sis00] Sistiaga, M., Navigation référencée images de terrain pour engins sous-marins, Ph.D Thesis, University of Montpellier II, September 2000.

[Tom91] Tomasi, C. and Kanade, T., Detection and Tracking of Point Features, Rapport technique CMU-CS-91-132, Carenegie Mellon University, April 1991.

[Tor93] Torr, P. H. S. and Murray, D. W., Outlier Detection and Motion Segmentation, In Proceedings SPIE Sensor Fusion VI, pp. 432-443, Boston, September 1993.