Uncalibrated Image Rectification for Coplanar Stereo Cameras

Vinicius Cesar[1], Thiago Farias[2], Saulo Pessoa[1], Samuel Macedo[1], Judith Kelner[1] and Ismael Santos[3]
[1]*Centro de Informatica, UFPE, Recife, Brazil*
[2]*Universidade de Pernambuco, Caruaru, Brazil*
[3]*Tecgraf, PUC-RIO, Rio de Janeiro, Brazil*

Keywords: Rectification, Stereo, Calibration, Reconstruction.

Abstract: Nowadays, underwater maintenance tasks, mostly in the case of oil and gas industries, have been assisted by computer vision algorithms. An important part of these procedures is the rectification of stereo images, which is the first step in the stereo 3D reconstruction pipeline. Some aspects of the underwater environment make the rectification process difficult: it presents a very noisy scenario; and the equipment is almost textureless. As a result of this demanding scenario, this article proposes a novel technique for a more accurate rectification of a set of images than the state-of-the-art methods. Tests were carried out proving the efficiency of the proposed technique.

1 INTRODUCTION

Cameras are broadly used by the industry to assist maintenance tasks, especially when the environment is unreachable or even harmful for human beings. By using cameras, an individual can remotely supervise the operation and take records for future consultation. Beyond these benefits, the captured footage can also be used by a computer vision application to provide additional information about the environment, such as its 3D structure. This task is usually performed by synchronized stereo rigs, which always require an image rectification stage in the application pipeline to reduce processing time and complexity.

The problem tackled by this work occurs in a deep underwater environment (depth can exceed $1000m$) where the maintenance task of a flexible pipe is carried out. Because of the high pressure of the environment, the task is performed by using ROVs (Remotely Operated Vehicles) equipped with a pair of cameras. Since no natural light can achieve such a depth, special low light cameras are used. Despite these cameras are sensitive to low light conditions (10^{-3} LUX), they are analog (resolution is limited to the NTSC standard) and can only capture gray scale images. In addition, the underwater environment is very noisy with particles floating around and the pipe is almost textureless. All these characteristics lead to a final poor quality image with limited contrast, which makes any feature extraction almost unfeasible. In order to overcome this problem, some high contrast markers were previously painted over the pipe. It consists in interleaved white and black regions along the pipe surface. So, by using these markers a tailored solution for the pipe segmentation was developed. This technique can provide a few stable set of features required by the rectification technique. The tests performed so far have shown that, at least for the presented study case, the segmentation solution using temporal coherence is robust enough to produce no outliers. However, it can be extended with RANSAC (Fischler and Bolles, 1981) in order to ensure robustness in more critical cases. A sample of the captured images and the segmentation results can be seen in Fig. 1. It worth to mention that this painting is required not

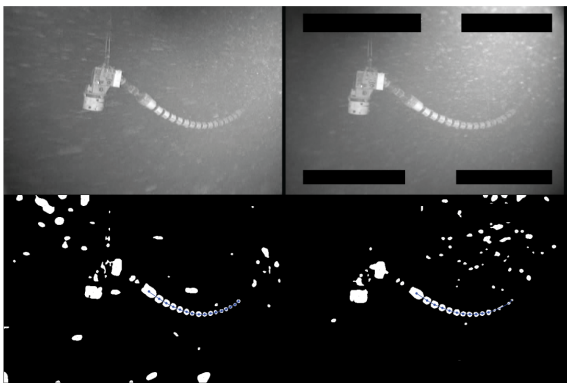


Figure 1: First row exhibits left and right images captured by the stereo rig.

Cesar V., Farias T., Pessoa S., Macedo S., Kelner J. and Santos I..
Uncalibrated Image Rectification for Coplanar Stereo Cameras.
DOI: 10.5220/0004745106480655
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 648-655
ISBN: 978-989-758-009-3

only by the feature extraction process, but by the subsequent tracking stage that will not be approached in this paper. Then, painting the pipe would be required even if the rectification process would not exist. Previously calibrating the pair of cameras is something that would be inconvenient because the cameras are mounted during the maintenance task and the technicians are not trained for it.

The current methods for rectification do not work properly in a noisy environment with a reduced number of feature correspondences, especially when the calibration is unknown. Thus, this paper presents a novel rectification technique for stereo rigs that operates even with a reduced number of feature correspondences. The state of the art techniques (such as (Fusiello and Irsara, 2008)) need at least six correspondences, while the proposed technique requires only three. However, the proposed solution requires the following restrictions on the cameras rig: cameras' projection planes must be coplanar; and cameras must have equal intrinsic parameters.

Tests were carried out in order to evaluate the proposed technique. Three different sets of tests were applied to measure the error related to the technique. The first one considered a synthetic test that was proposed only with points numerically disturbed by a random generated noise. In the second test, a real structured environment was built and a carefully mounted stereo rig was used.

The third test occurred in a real scenario. As stated before, the technique was tested in a deep underwater environment, where images were captured by a ROV.

2 RELATED WORK

Rectification of stereo images is a frequently investigated topic by the computer vision community. These researches began by photogrammetrists, such as (Slama et al., 1980), which were further developed by computer vision researchers aiming to facilitate the feature matching between images from a stereo rig.

Rectification techniques can be classified into two categories: calibrated, and uncalibrated. Calibrated techniques assume that cameras' intrinsic and extrinsic parameters are known and the rectifying homographies are estimated only by taking into account these parameters. In (Fusiello et al., 2000), a simple and effective calibrated method is presented.

Uncalibrated techniques estimate the rectifying homographies by using a set of corresponding 2D points between the images and/or epipolar restrictions (such as the fundamental matrix). These techniques are more used than the calibrated ones because in most of the real problems the rectification is required in a stage before the cameras poses are known. However, it is a more complex problem with non-linear solutions, which requires the use of approximations and optimization methods. Such methods are required because there are infinite pairs of rectifying homographies, although it is convenient to choose the one which produces less image deformation. Some uncalibrated techniques can be found in (Hartley, 1998), (Loop and Zhang, 1999), (Isgro and Trucco, 1999), (Fusiello and Irsara, 2008).

When the epipole is close to or inside the image, image deformation tends to be large. In these cases planar rectifications are not enough, therefore it is necessary to use different techniques such as cylindrical rectification (Roy et al., 1997) or polar rectification (Pollefeys et al., 1999).

In the scenario presented by this paper only part of cameras parameters are previously known, which enforces the use of an uncalibrated technique. However, uncalibrated techniques require at least six accurate corresponding points between the images (Fusiello and Irsara, 2008), requirement that may not always be fulfilled by the application. The proposed technique overcomes this limitation by relying on some restrictions imposed on the stereo rig. In practice, these constrain the way in which cameras must be relatively positioned and oriented. If the stereo rig is mounted so that the cameras' projection planes are coplanar, epipoles will be localized close to the infinity, enabling a planar rectification to solve the problem.

3 BACKGROUND

In this section, some concepts that are at the core of the proposed rectification technique will be presented as well as the adopted notation. These concepts are more extensively explained in (Hartley and Zisserman, 2004), (Loop and Zhang, 1999).

3.1 Epipolar Geometry

Given two pinhole cameras $\mathcal{P}$ and $\mathcal{P}'$ with their respective projection matrices defined as $\mathtt{P} = \mathtt{K}[\mathtt{I}|\mathbf{0}]$ and $\mathtt{P}' = \mathtt{K}'\mathtt{R}[\mathtt{I}|-\mathbf{C}]$. $\mathtt{I}$ is a 3×3 identity matrix. Camera $\mathcal{P}$ has its projection center at the origin of the coordinate system $\mathbf{0} = [0,0,0]^\top$. Camera $\mathcal{P}'$ has its projection center at $\mathbf{C} = [x_c, y_c, z_c]^\top$, defined in Euclidean coordinates. Furthermore, matrices $\mathtt{K}$ and $\mathtt{K}'$ are the so called calibration matrices, which encapsulate cameras' intrinsic parameters. A simplified calibration matrix has the form $diag(f, f, 1)$, where f is the lens

focal length. This simplified form assumes that the principal point is the center of the image where the pixel skew is zero and the pixel aspect ratio is one.

The canonical form of the cameras matrices $\mathtt{P}$ and $\mathtt{P}'$ are calculated aplying a projective transformation to the 3D space such that $\mathtt{P} = [\mathtt{I}|\mathbf{0}]$.

Given a 3D point $\mathbf{X}$ in homogeneous coordinates, its projection in image I through camera $\mathcal{P}$ is given by $\mathbf{x} = \mathtt{P}\mathbf{X}$. Likewise, $\mathbf{x}' = \mathtt{P}'\mathbf{X}$ is the projection of $\mathbf{X}$ in image I' through the camera $\mathcal{P}'$.

By using these projected points, the epipolar constraint can be established as

$$\mathbf{x}'^{\top}\mathtt{F}\mathbf{x} = 0, \tag{1}$$

which is valid for all 2D point correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$. $\mathtt{F}$ is a 3×3 rank 2 matrix named fundamental matrix, which maps points from image I to lines (named epipolar lines) in image I'. Given the line $\mathbf{l}' = \mathtt{F}\mathbf{x}$ in image I', it can be said that $\mathbf{x}'$ lies on $\mathbf{l}'$ since $\mathbf{x}'^{\top}\mathbf{l}' = 0$. The reverse idea is also valid, and therefore point $\mathbf{x}$ lies on the line $\mathbf{l} = \mathtt{F}^{\top}\mathbf{x}'$. All epipolar lines of one image intersect each other at a single point named epipole, where $\mathbf{e}$ is the epipole in I and $\mathbf{e}'$ is the epipole in I'.

Being $\mathtt{P}_{can}$ and $\mathtt{P}'_{can}$ two projection matrices from canonical cameras, *i.e.* $\mathtt{P}_{can} = [\mathtt{I}|\mathbf{0}]$ and $\mathtt{P}'_{can} = \mathtt{P}'\mathtt{H}_{can} = [\mathtt{M}|\mathbf{m}]$, the fundamental matrix between the two images captured by these cameras can be defined as

$$\mathtt{F} = [\mathbf{m}]_{\times}\mathtt{M}, \tag{2}$$

where $[\mathbf{m}]_{\times}$ stands for the antisymmetric matrix that is equivalent to the cross product with $\mathbf{m}$.

Two images $\bar{I}$ and $\bar{I}'$ are said to be rectified if all matching points $\bar{\mathbf{x}} = [\bar{x}, \bar{y}, 1]^{\top}$ and $\bar{\mathbf{x}}' = [\bar{x}', \bar{y}', 1]^{\top}$ have the same coordinate in y, *i.e.* $\bar{y} = \bar{y}'$. Thus, with the rectified matching points on the same line, the stereo matching is made easier and computationally faster.

The rectifying process consists in estimating two homographies $\mathtt{H}$ and $\mathtt{H}'$, which when applied to images I and I', respectively, make them rectified. The epipolar geometry between two rectified images has some noteworthy particularities. The fundamental matrix between two rectified images is $\bar{\mathtt{F}} = [[1,0,0]^{\top}]_{\times}$

All the epipolar lines of a rectified image are parallel to the x direction of the image, since $\bar{\mathbf{l}}' = \bar{\mathtt{F}}\bar{\mathbf{x}} = [0,1,-\bar{y}]^{\top}$. Assuming all epipolar lines intersect at the epipoles, the epipoles are valued $[1,0,0]^{\top}$.

3.2 Loop and Zhang Algorithm

Loop and Zhang in (Loop and Zhang, 1999) present a rectification algorithm that uses the epipolar restrictions of the images. This algorithm aims to rectify images by minimizing the distortion caused by projective transformations. The algorithm requires the fundamental matrix and the epipole in the first image.

The strategy adopted by the algorithm is to decompose the rectifying homographies in three transformation: 1) a projective transformation $\mathtt{H}_p$, that maps the epipoles to the infinity; 2) a similarity transformation $\mathtt{H}_r$, that rotates and translates the epipoles to $[1,0,0]^{\top}$; and 3) a shearing transformation $\mathtt{H}_s$ that minimizes image distortion in the x coordinates. Using the notation defined in Section 3.1, the transformations $\mathtt{H}$ and $\mathtt{H}'$, which rectify the images I and I' respectively, are defined as

$$\mathtt{H} = \mathtt{H}_s\mathtt{H}_r\mathtt{H}_p \tag{3}$$

and

$$\mathtt{H}' = \mathtt{H}'_s\mathtt{H}'_r\mathtt{H}'_p. \tag{4}$$

To compute the projective transformation, two lines must be defined: $\mathbf{w} = [w_1, w_2, w_3]^{\top}$ and $\mathbf{w}' = [w'_1, w'_2, w'_3]^{\top}$. The lines $\mathbf{w}$ and $\mathbf{w}'$ pass through epipoles $\mathbf{e}$ and $\mathbf{e}'$, respectively. In order to map the epipoles to infinity, one has to define projective the transformations $\mathtt{H}_p$ and $\mathtt{H}'_p$ that respectively map $\mathbf{w}$ and $\mathbf{w}'$ to infinity. Since there are an infinity number of possible lines, it is preferred to choose the ones that minimize image distortions. Therefore, the projective transformations are defined as

$$\mathtt{H}_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ w_1 & w_2 & w_3 \end{bmatrix}. \tag{5}$$

Similarly we can define $\mathtt{H}'_p$.

After projective transformations, epipolar lines become parallel one another considering the same image, although they are not aligned considering the matching lines between the images. The similarity transformations rotate and translate images in order to make the epipolar lines parallel to the x direction. These transformations are

$$\mathtt{H}_r = \begin{bmatrix} F_{32} - w_2F_{33} & w_1F_{33} - F_{31} & 0 \\ F_{31} - w_1F_{33} & F_{32} - w_2F_{33} & F_{33} + v'_c \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

and

$$\mathtt{H}'_r = \begin{bmatrix} w'_2F_{33} - F_{23} & F_{13} - w'_1F_{33} & 0 \\ w'_1F_{33} - F_{13} & w'_2F_{33} - F_{23} & v'_c \\ 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

where v'_c is a common vertical translation for both images.

The homographies $\mathtt{H}_r\mathtt{H}_p$ and $\mathtt{H}'_r\mathtt{H}'_p$ are already able to rectify the images, although shearing transformations can be added in order to minimize images distortion. This transformation only modify x coordinates, without affecting the rectification. In short, it is simply an attempt to preserve perpendicularity and aspect ratio of the images.

4 METHODOLOGY

The following methodology draws its actions from the constraints aforementioned, where the cameras have coplanar projection planes and the identical intrinsic parameters. Since camera $\mathcal{P}$ is at the origin, its projection matrix can be expressed as $\mathtt{P} = \mathtt{K}[\mathtt{I}|\mathbf{0}]$, where $\mathtt{K}$ is the calibration matrix (intrinsic parameters). Matrix $\mathtt{K}$ is stated as $diag(f, f, 1)$, where f is the lens focal length. Camera $\mathcal{P}'$ has a translation along the xy plane and a rotation by θ around its optical axis. In addition, camera $\mathcal{P}'$ has the same intrinsic calibration of camera $\mathcal{P}$. Thus, the projection matrix of camera $\mathcal{P}'$ is defined as $\mathtt{P}' = \mathtt{KR}[\mathtt{I}|-\mathbf{C}]$, where $\mathtt{R}$ is a tridimensional counterclockwise rotation by angle θ about z axis and $\mathbf{C} = [x_c, y_c, 0]^\top$. In order to simplify further calculations, $\mathbf{C}$ vector will be presented as $\mathbf{C} = d[\cos\alpha, \sin\alpha, 0]$, with $d = ||\mathbf{C}||$.

To obtain the canonical form of the camera matrices, we can define the transformation

$$\mathtt{H}_{can} = \begin{bmatrix} \mathtt{K}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \tag{8}$$

resulting in $\mathtt{P}_{can} = [\mathtt{I}|\mathbf{0}]$ and $\mathtt{P}'_{can} = \mathtt{P}'\mathtt{H}_{can} = [\mathtt{KRK}^{-1}| - \mathtt{R}\mathbf{C}] = [\mathtt{R}| - \mathtt{R}\mathbf{C}]$. By using these matrices and (2) one can calculate the fundamental matrix related to $\mathcal{P}$ and $\mathcal{P}'$, resulting in

$$\begin{aligned} \mathtt{F} &= [-\mathtt{R}\mathbf{C}]_\times \mathtt{R} \\ &= -d \begin{bmatrix} 0 & 0 & \sin(\alpha+\theta) \\ 0 & 0 & -\cos(\alpha+\theta) \\ -\sin\alpha & \cos\alpha & 0 \end{bmatrix}. \end{aligned} \tag{9}$$

Once the fundamental matrix is up to scale, factor $-d$ can then be removed from (9). The epipoles from $\mathtt{F}$ and $\mathtt{F}^\top$ are extracted using the nullspace of these matrices, giving respectively

$$\mathbf{e} = null(\mathtt{F}) = [\cot\alpha, 1, 0]^\top \tag{10}$$

and

$$\mathbf{e}' = null(\mathtt{F}') = [\cot(\alpha+\theta), 1, 0]^\top. \tag{11}$$

After calculating the epipolar geometry, the rectifying homographies can be found by applying the Loop and Zhang's algorithm (Loop and Zhang, 1999).

The first step is to define the projective transformations that map epipoles to infinity. In order to define these transformations, one has to determine the lines $\mathbf{w}$ and $\mathbf{w}'$.

As stated in (10) and (11), epipoles are already at infinity if cameras are coplanar, so the line at infinity $l_\infty = [0, 0, 1]^\top$ must be chosen in order to avoid image distortion. Then

$$\mathbf{w} = \mathbf{w}' = [0, 0, 1]^\top, \tag{12}$$

which, by (5), leads to $\mathtt{H}_p = \mathtt{H}'_p = \mathtt{I}$.

The following step maps, through a rotation and translation, the epipoles onto the point $[1, 0, 0]^\top$. In (Loop and Zhang, 1999), the mapping is given by (6) and (7), which depends on $\mathtt{F}$, $\mathbf{w}$, and $\mathbf{w}'$. Using (9) and (12) to fill (6) and (7), one can get

$$\mathtt{H}_r = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

and

$$\mathtt{H}'_r = \begin{bmatrix} \cos(\theta+\alpha) & \sin(\theta+\alpha) & 0 \\ -\sin(\theta+\alpha) & \cos(\theta+\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

The last step of Loop and Zhang's algorithm determines affine transformations in order to preserve the aspect ratio and perpendicularity of image. It is also worth to mention that $\mathtt{H}_p\mathtt{H}_r$ and $\mathtt{H}'_p\mathtt{H}'_r$ are rigid transformations, therefore this step is not necessary. So, one can define $\mathtt{H}_s = \mathtt{H}'_s = \mathtt{I}$.

By using the rectifying homographies calculated as (3) and (4), the proposed method, adapted from Loop and Zhang's technique for cameras with coplanar projection planes can be summarized as follows. Given two images I and I' and 2D matching points $\mathbf{x} \leftrightarrow \mathbf{x}'$, where $\mathbf{x}$ is in I and $\mathbf{x}'$ is in I', the rectification of I and I' consists in calculating the angle α, angle $\beta = \alpha + \theta$ and matches $\bar{\mathbf{x}} \leftrightarrow \bar{\mathbf{x}}'$, where $\bar{\mathbf{x}} = R(\alpha)\mathbf{x}$, $\bar{\mathbf{x}}' = R(\beta)\mathbf{x}'$, and $R(\theta)$ is a matrix representing a 2D clockwise rotation by angle θ.

According Loop and Zhang in (Loop and Zhang, 1999), the rectification is done from the fundamental matrix, although one can define another approach using 2D point matches to determine angles α and β. Given

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} x\cos\alpha - y\sin\alpha \\ x\sin\alpha + y\cos\alpha \end{bmatrix} \tag{15}$$

and

$$\bar{\mathbf{x}}' = \begin{bmatrix} \bar{x}' \\ \bar{y}' \end{bmatrix} = \begin{bmatrix} x'\cos\beta - y'\sin\beta \\ x'\sin\beta + y'\cos\beta \end{bmatrix}, \tag{16}$$

and knowing that the rectified image obeys constraint $\bar{y} = \bar{y}'$, one can have

$$x\sin\alpha + y\cos\alpha - x'\sin\beta - y'\cos\beta = 0. \tag{17}$$

If there are n 2D matching points, there will be n equations like (17), which leads to the system

$$\begin{bmatrix} x_1 & y_1 & -x'_1 & -y'_1 \\ x_2 & y_2 & -x'_2 & -y'_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & -x'_n & -y'_n \end{bmatrix} \begin{bmatrix} \sin\alpha \\ \cos\alpha \\ \sin\beta \\ \cos\beta \end{bmatrix} = \mathtt{A}\mathbf{y} = \mathbf{0}, \tag{18}$$

where $\mathbf{0}$ is a column n-vector of zeros.

The solution has two degrees of freedom and can be reached by using only 2D matches between the two images. The degrees of freedom are the two angles that rotate the respective images. The aforementioned system is non-linear due to sine and cosine functions, and a linear approximation is not suited for a noisy scenario.

The solution of the problem given in (18) can be found by using two different methods that will be described in the next subsections.

4.1 Linear Solution

This approach uses a technique called "linearization" that was employed by (Ansar and Daniilidis, 2003) and (Lepetit et al., 2009) to estimate pose of cameras based on correspondences between 2D and 3D points. This technique modifies the presentation of the problem to apply a linear approximation that satisfies its non-linear constraints.

In order to make the problem linear, one can substitute the non-linear part of the problem by new variables, giving $\mathbf{y} = [\sin\alpha, \cos\alpha, \sin\beta, \cos\beta]^\top = [y_1, y_2, y_3, y_4]^\top$. One must ensure that the Pythagorirean trigonometric identities

$$y_1^2 + y_2^2 = 1 \tag{19}$$

and

$$y_3^2 + y_4^2 = 1 \tag{20}$$

still hold.

In order to find the solution space of (18) one can solve it by using a SVD decomposition $\mathtt{A} = \mathtt{UDV}^\top$. The approximated solution is within the space defined by the base composed of the third and fourth columns of $\mathtt{V}$, named $\mathbf{u}$ and $\mathbf{v}$, respectively, which are related to the smallest singular values. Since there are two variables, the solution space is two-dimensional. Thus, the solution to $\mathbf{y}$ must be a linear combination of the vectors $\mathbf{u} = (u_1, u_2, u_3, u_4)$ and $\mathbf{v} = (v_1, v_2, v_3, v_4)$, resulting

$$\mathbf{y} = \gamma\mathbf{u} + \delta\mathbf{v}. \tag{21}$$

By replacing (21) in (19) and (20), one can have

$$(\gamma u_1 + \delta v_1)^2 + (\gamma u_2 + \delta v_2)^2 = 1 \tag{22}$$

and

$$(\gamma u_3 + \delta v_3)^2 + (\gamma u_4 + \delta v_4)^2 = 1, \tag{23}$$

which represent two ellipses, because $\Delta_1 = -2(v_1u_2 - v_2u_1)^2$ and $\Delta_2 = -2(v_3u_4 - v_4u_3)^2$ are always negative.

Adding up (22) and (23), one can find

$$\begin{aligned}\mathbf{u}^\top\mathbf{u}\gamma^2 + \mathbf{u}^\top\mathbf{v}\gamma\delta + \mathbf{v}^\top\mathbf{v}\delta^2 = \\ \gamma^2 + \delta^2 = 2,\end{aligned} \tag{24}$$

once $\mathbf{u}$ and $\mathbf{v}$ were picked from an orthonormal basis. Conic in (24) describes a circumference whose equation is satisfied by intersection points of ellipses (22) and (23). In this case, there can be two or four intersections.

The intersections of the circumference with the two ellipses can be found by solving a fourth order polynomial such as $ax^4 + bx^2 + c = 0$, which has two symmetric solutions. Such polynomial can be determined by using the Sylvester resultant. The result can be used in (24) to determine the last variable. The symmetric solutions are realistic because the supplement of an answer is also a correct answer, once images remain rectified when rotated by 180°. In the case where there are four solutions, the one that satisfies $\gamma > \delta$ must be used.

4.2 Non-linear Solution

The system given by (18) is non-linear, and thus not suitable for a direct solution. Therefore, another approach to tackle the problem is by using numerical methods. Finding α and β that minimize $||\mathtt{A}\mathbf{y}||^2$ is a least squares problem that can be solved by the Gauss-Newton method. The initial values of α and β, namely α_0 and β_0, can be assigned in two different ways: either by using the result of the linear method described in this paper, or by taking both values as zero. The second choice is acceptable because the cameras are mounted on ROVs manually attempting to enforce the coplanar constraints.

5 RESULTS

In an attempt to evaluate the proposed technique, this section proposes three sets of tests, which respectively compare the approaches proposed one another with synthetic data, compare the best approach with the state-of-the-art technique described by Fusiello and Irsara (Fusiello and Irsara, 2008), and show the epipolar error in a real cluttered environment.

5.1 Synthetic Simulation

The simulation proposes to compare the different approaches of the suggested method to solve the homogeneous system given by (18) minimizing $||\mathtt{A}\mathbf{y}||$. A synthetic scene was created with two cameras, $\mathcal{P}$ and $\mathcal{P}'$, with coplanar projection planes. The intrinsic calibration of the cameras were generated considering an image size of 800×600 pixels and a focal length of 965.68 pixels (a horizontal field of view around 45°).

Table 1: Synthetic simulation results with 2 pixels of Gaussian noise (Mean±std and Iterations).

	6 points			12 points			20 points		
	α-Error	β-Error	Itr	α-Error	β-Error	Itr	α-Error	β-Error	Itr
Linear	2.81 ± 2.91	1.71 ± 1.83	–	1.88 ± 1.91	1.16 ± 1.17	–	1.45 ± 1.49	0.90 ± 0.91	–
Non-Linear	1.16 ± 2.00	0.62 ± 1.21	5.64	0.68 ± 0.56	0.37 ± 0.32	5.14	0.50 ± 0.41	0.28 ± 0.24	4.96
Linear+Non-Linear	1.11 ± 1.14	0.59 ± 0.64	4.29	0.68 ± 0.56	0.37 ± 0.32	3.83	0.50 ± 0.41	0.28 ± 0.24	3.59

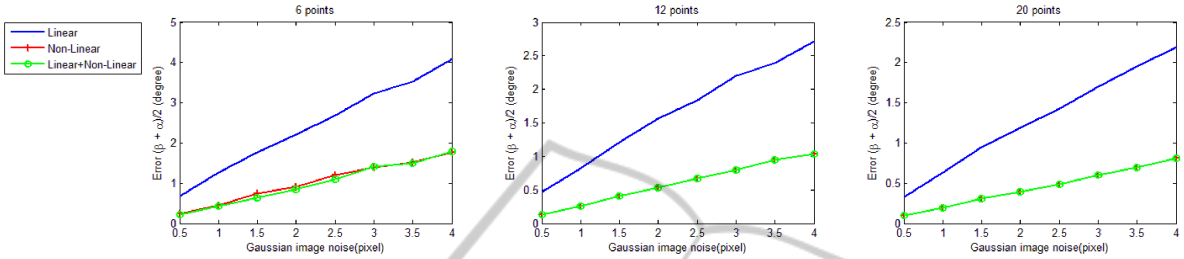


Figure 2: Synthetic results varying the amount of point correspondences and Gaussian noise.

In this simulation the cameras are 90*cm* apart from each other and a 3D point cloud was randomly generated inside the frustums of the cameras and away between 3*m* and 6*m* from their baseline. These numbers represent the expected configuration of the real environment where the method will be applied. The 3D points are projected by the cameras $\mathcal{P}$ e $\mathcal{P}'$ and thereafter a Gaussian noise will be added to the projections to simulate the lack of precision of the tracking process. These projected points are the input for the rectification method proposed in this work. The non-linear technique will be evaluated against both initializing α_0 and β_0 with zero values and with the solution of the linear approach.

The results were obtained using 6, 12 and 20 3D points. The Gaussian noise applied to the projections has standard deviation from 0.5 to 4 pixels. For each noise (standard deviation) generated, the tests were computed 2000 times in order to reach an accurate evaluation. In each sample, the values used for α and θ came from a uniform distribution generating values between -30° and 30°. The results are illustrated in Fig. 2. Table 1 presents the numeric results.

Note that the precision is strongly related to the number of points. It is also possible to observe that the linear algorithm achieved the worst performance with errors greater than 4°, while the non-linear achieved averaging errors below 2°.

When the result of the linear algorithm feeds the initial values of the non-linear algorithm, which uses an iterative Gauss-Newton algorithm, one can observe that the results are very similar to the ones where the initial values are taken from zero ($\alpha_0 = 0$ and $\beta_0 = 0$). The difference between both approaches does not exceed 0.1°. This occurs because the local minima do not influence the convergence of the algorithm, except when the amount of points is small.

In Table 1, one can verify that initializing the Gauss-Newton algorithm with the linear approach decreases between 20% and 40% the number of iterations. It is possible to observe as well that the convergences in the non-linear approaches are different only when it has six points.

Overall, one can conclude that an optimal solution to the problem can be reached applying the Gauss-Newton method, unless the amount of points is small (around 6). In this case, by using as initial value the output of the linear approach can produce more accurate results.

All simulations were performed using MATLAB. The hardware used in the tests was a computer with an Intel Core i7 3960X 3.30Ghz processor and 24GB RAM. The execution time was collected using a simulation with 20 3D points. In average, the linear approach takes around 0.15*ms* to finish, the non-linear 0.3*ms*, and the linear+non-linear also 0.3*ms*. The reason why the two last approaches spend the same time is because the non-linear approach performs fewer iterations when initialized by the linear approach result.

5.2 Controlled Environment Tests

In this test, a controlled environment is used to compare the proposed approach to the state-of-the-art technique of Fusiello and Irsara (Fusiello and Irsara, 2008). The test consists of two pictures of a chessboard, as illustrated in Fig. 3. The camera used in this test was a Canon T4i and the picture resolution was set to 720×480 pixels. Only this resolution was tested because it is closest one to the resolution of the cameras attached to the ROV. The camera was positioned about 70*cm* away from the target. Between the two shots, the camera was moved 20*cm* rightward and 4*cm* upward keeping the same focal length and en-

Figure 3: Images used for the controlled environment tests.

forcing coplanarity constraint of cameras. The camera was intentionally moved without an accurate process (actually it was manually moved), while keeping the optical axes nearly parallel. Also, the images were manually defocused to simulate the blur phenomenon that occurs underwater.

Since the exact position of camera is unknown, there is no ground truth. Thus, the technique will be evaluated using the rectification error (*i.e.* epipolar error, or the distance of the point to the related epipolar line, which can be calculated as $\mathbf{Ay}$). The chessboard has 54 points that can be extracted and matched between the images. To apply the rectification, a subset ranging from 6 to 20 points chosen randomly will be used as input, although all 54 points are used to measure the epipolar error. It is expected that the more precise the rectification the smaller will be the errors. For each amount of points, 100 subsets of random points were chosen to be rectified with the nonlinear approach proposed (initialized with the linear approach) and later with the technique proposed by Fusiello and Irsara (Fusiello and Irsara, 2008).

In Fig. 4, it can be seen that the Fusiello and Irsara's technique had poor results with a small amount of points, since such technique has more degrees of freedom to be determined. However, from 12 points on, the Fusiello and Irsara's technique can estimate more precisely all the system variables than the proposed technique.

In Fig. 5 the rectification of the left image using 6 points is shown. The rectification from Fusiello and Irsara applied more distortion to the images and has a strong projective distortion, as well. Such result is not acceptable as the epipoles are close to the infinity. As expected, the proposed technique applied a simple rotation.

The tests were carried out using the same hardware from the synthetic simulation. The proposed technique, due to its complex minimization calculations, had an execution time between 0.2 and 0.3*ms*. Fusiello and Irsara's technique had an average time of 230*ms* with 6 points and 3*s* with 20 points.

5.3 Real Experiments

This work was also tested with real underwater images. The cameras used for the operations were Kongsberg OE15-100c (low light cameras for high depths). These cameras are attached to the ROV by a metal support and are 45*cm* apart. The support of the cameras is an attempt to keep the cameras' projection planes parallel. The system first segments the flexible pipe from both left and right images in order to extract features. The segmentation is performed in two stages: first, a tailored thresholding technique is used to find out which regions are potentially of the pipe (the white blobs in the second row of Fig. 1); second, a search is performed in order to discover the best sequence of thresholded regions which describe a pipe. Since the features are extracted by evaluating the centroid of the thresholded regions, features position are not very precise. These features are then matched between both images. The amount of extracted features ranges from 5 to 16. Even without a groundtruth, the achieved results can testify the technique efficiency.

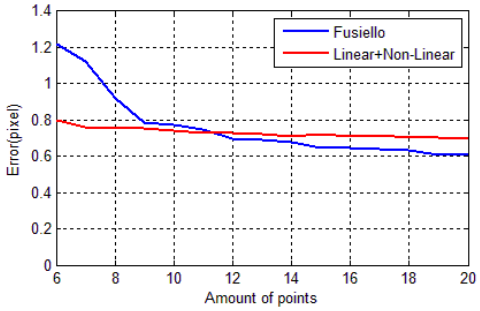
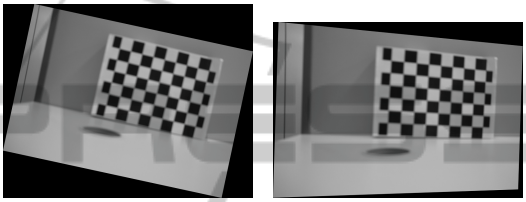


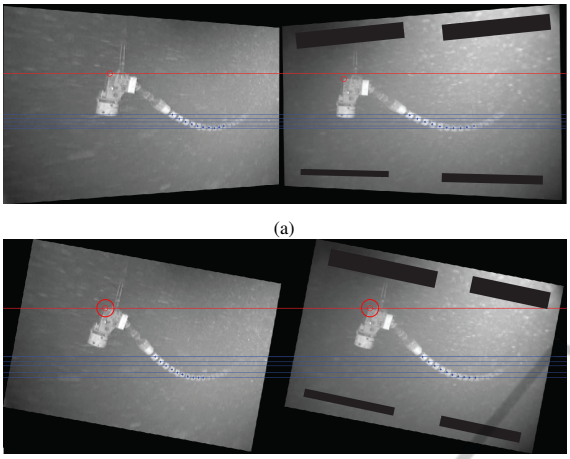
Figure 4: Comparison of the results obtained with Fig. 3 by the proposed technique and the Fusiello and Irsara's technique.



(a) (b)

Figure 5: Rectification of the left image of Fig. 3 using (a) the proposed technique and (b) Fusiello and Irsara's technique.

Fusiello and Irsara's technique was also tested with the real underwater images. However, it fails in many cases because the estimated homographies produce huge projective distortions, while only a rotation and a small projective transformation are needed (as illustrated in Fig. 6(a)). In addition, consecutive pairs of images (*i.e.*, similar images) produced completely different homographies, which is mostly due to inaccuracies in the detection of features position. In Fig.



(a)

(b)

Figure 6: Rectification of the images of Fig. 1 using (a) Fusiello and Irsara's and (b) the proposed technique.

6(a) the red circle shows that this part of the image is not correctly rectified.

The proposed technique achieved better results than the Fusiello and Irsara's technique. The results along multiple frames were also more stable. The rotation angle in the first image was between 9° and 13° while the second image was between 10° and 14°. The epipolar error ranges from 0.5 to 1.2 pixels. Fig. 6(b) illustrate the result. The red circle shows that, in contrast with the Fusiello and Irsara's technique, this part of the image is correctly rectified.

After the initial rectification, the images were aligned allowing the extraction of more information about the visual landmarks in order to perform a more accurate 3D reconstruction. Thus, the error embedded in the rectification process is acceptable for the whole system. Outliers were not detected in the tests, however they could be removed using RANSAC-based algorithms.

6 CONCLUSIONS

This paper proposed a novel rectifying technique for images under constraints imposed by underwater maintenance tasks. The proposed technique takes advantage of the geometry of the structure of the stereo rig, which is positioned keeping the cameras' projection planes coplanar. This arrangement represents lesser degrees of freedom for the rectification problem, which allows lesser point correspondences to obtain satisfactory accuracy, as well.

Tests were carried out using synthetic data, a real controlled environment and a real underwater scene. The proposed technique performed better than the state-of-the-art method (Fusiello and Irsara, 2008). As future work, the present technique will be improved to take into consideration variations on the intrinsic parameters of the cameras, such as focal length and principal point, which were considered to be fixed under the performed tests.

REFERENCES

Ansar, A. and Daniilidis, K. (2003). Linear pose estimation from points or lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:282–296.

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

Fusiello, A. and Irsara, L. (2008). Quasi-euclidean uncalibrated epipolar rectification. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4.

Fusiello, A., Trucco, E., and Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Mach. Vision Appl.*, 12(1):16–22.

Hartley, R. I. (1998). Theory and practice of projective rectification.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Isgro, F. and Trucco, E. (1999). Projective rectification without epipolar geometry. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages –99 Vol. 1.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o(n) solution to the pnp problem. *Int. J. Comput. Vision*, 81(2):155–166.

Loop, C. and Zhang, Z. (1999). Computing rectifying homographies for stereo vision. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 2 vol. (xxiii+637+663).

Pollefeys, M., Koch, R., and Van Gool, L. (1999). A simple and efficient rectification method for general motion. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 496 –501 vol.1.

Roy, S., Meunier, J., and Cox, I. (1997). Cylindrical rectification to minimize epipolar distortion. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 393 –399.

Slama, C. C., Theurer, C., and Henriksen, S. W., editors (1980). *Manual of Photogrammetry*. American Society of Photogrammetry.