# Stereo Vision System for Real Time Inspection and 3D Reconstruction

Lenildo C. SILVA[1], Antonio PETRAGLIA[2], Senior Member, IEEE, Mariane R. PETRAGLIA[3], Member, IEEE

[1]PEE/COPPE, UFRJ, Cidade Universitária, Rio de Janeiro, Brazil, e-mail : lenildo@pads.ufrj.br
[2]PEE/COPPE, UFRJ, Cidade Universitária, Rio de Janeiro, Brazil, e-mail : petra@pads.ufrj.br
[3]PEE/COPPE, UFRJ, Cidade Universitária, Rio de Janeiro, Brazil, e-mail : mariane@pads.ufrj.br

***Abstract*— This work presents a three-dimensional vision system for inspection activities of installations by remotely operated vehicles. A real-time stereo vision system is used for the acquisition of stereo pairs of images that, after preprocessing, are submitted to a reconstruction procedure in order to obtain three-dimensional coordinates, to perform dimensioning of objects in the acquired images.**

***Index Terms*— photogrammetry, reconstruction, robot vision, stereo vision**

## I. Introduction

This work reports a stereo vision system suitable for the activities of visual inspection in real time. Its methodology consists of suppressing objects that are not of interest from the images of the structure being evaluated, by carrying out a preprocessing of these images to obtain with greater reliability the parameters to be extracted. Through techniques of analytic photogrammetry, a reconstruction procedure is applied to the stereo images in order to obtain 3-D coordinates from the scene, and then to perform the dimensioning of objects of interest. Based on this methodology, a modular, integrated (software and hardware) three-dimensional vision system has been developed, which performs the acquisition and processing of stereo pairs [1].

The three-dimensional vision system can be divided into three stages, as illustrated in Fig. 1. The acquisition stage consists of a real-time stereo image acquisition consisting of a pair of video cameras connected to digitizer cards on a personal computer. Images of interest are digitized and recorded in stereo pairs. The preprocessing stage consists of improving the acquired images in order to adequate them to the reconstruction procedure, with the purpose of enhancing the accuracy of the parameters to be extracted from the images. Finally, the reconstruction stage consists on the estimation of 3-D coordinates from the stereo images, and then to perform the dimensioning of objects in the 3-D scene.

## II. Stereo Image Acquisition

The acquisition of the stereo images is made by a stereo vision system suited to submarine activities, to make it possible the visualization in real time of the scenes to be investigated [2]. The system is based on a personal computer, and uses two CCD cameras to the acquisition of the images. Its architecture is presented in Fig. 2.


Fig. 1. Block diagram of the real-time stereo system.

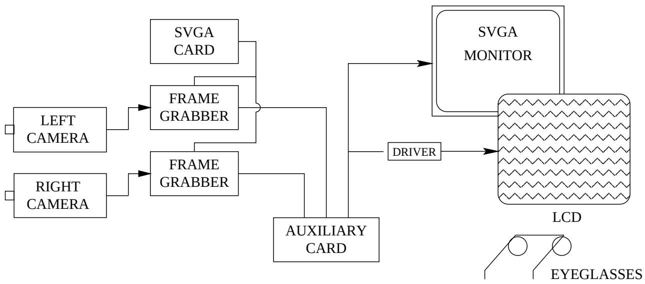



Fig. 2. Architecture of the stereo image acquisition system.

The CCD cameras have been assembled in a fixed basis, lined up horizontally, being the baseline corresponding to about 5.5 centimeters, (this value was chosen for visualization purposes). The digitizer cards are frame grabbers that convert the NTSC video signals coming from the cameras so that they can be shown on the SVGA monitor. The output of both cards are connected to an auxiliary board, whose function is to perform the switching between the left and right images alternatively, in a rate of 120 Hz, chosen to assure that each image is shown for a human eye in 1/60 seconds; it eliminates the blinking effect that would occur in showing the two images alternatively. The stereoscopic imaging device is composed by a LCD screen and passive eyeglasses, that allows the visualization of 3-D scenes from the composition of the 2-D images (left and right), making it possible the notion of depth.

## III. Preprocessing

The preprocessing procedure includes the application to the images previously acquired of some techniques of image processing such as histogram equalization and filtering. Among the filtering techniques used are low-pass filtering, used for noise smoothing and interpolation, and band-pass filtering used in the enhancement of edges [3]. The use of the techniques of edge detection allows a better characterization of the boundaries of each image, which is useful for the determination of the 2-D coordinates. Usually, a gradient operator is used to find the edges, following the application of a threshold to generating a binary edge map [4].

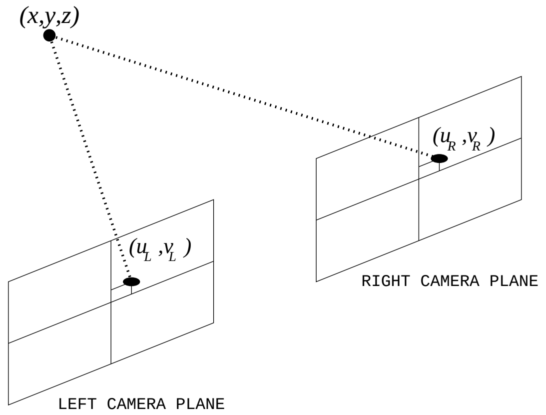



Fig. 3. 3-D point in space and its 2-D projections

## IV. RECONSTRUCTION

After the acquisition and preprocessing stages, the stereo images are submitted to the reconstruction stage, for estimating 3-D coordinates in the analyzed images, with the purpose of performing the dimensioning between points of interest in the scene. The reconstruction procedure is made based in photogrammetric methods.

Analytical photogrammetry includes a set of techniques by which, from measurements of one or more two-dimensional perspective projections of a 3-D object, one can make inferences about the 3-D position, orientation and dimensions of the observed 3-D object in a world reference frame [5], [6]. Fig. 3 illustrates the camera and world reference frames, showing a point in the 3-D space and its projection in the 2-D planes of each camera.

The complete specification of the orientation of a camera is given by the exterior orientation and the interior orientation. The exterior orientation refers to the computation of the parameters that determine the pose of the camera in the world reference frame. The interior orientation refers to the parameters that determine the geometry of a bundle of 3-D rays from the measured image coordinates, relating the geometry of the ideal perspective projection to the physics of the camera.

The relation between the camera reference frame and the world reference frame is given by a translation and a rotation. A point $\mathbf{x} = [\mathbf{x}\ \mathbf{y}\ \mathbf{z}]^{\mathrm{T}}$ in the world reference frame is expressed with respect to the camera reference frame through a translation given by the vector $\mathbf{t} = [\mathbf{x_0}\ \mathbf{y_0}\ \mathbf{z_0}]^{\mathrm{T}}$.

The perspective projection of the camera is obtained with respect to the $z$-axis, that is the optical axis of the camera. However, the directions of the $x-$, $y-$ and $z-$axes of the camera reference frame can differ from the directions of the world reference frame. The rotation by which the world reference frame is positioned into correspondence with the camera reference frame is a sequence of three rotations, expressed by a rotation matrix $\mathbf{R}(\omega, \phi, \kappa)$.

A point $\mathbf{x} = [\mathbf{x}\ \mathbf{y}\ \mathbf{z}]^{\mathrm{T}}$ of the world reference frame is represented by the point $\mathbf{x}' = [\mathbf{p}\ \mathbf{q}\ \mathbf{s}]^{\mathrm{T}}$ of the camera reference frame, where

$$\begin{bmatrix} p \\ q \\ s \end{bmatrix} = \mathbf{R}(\omega, \phi, \kappa) \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} \quad (1)$$

From this representation of the 3-D point in the camera reference frame, one can obtain its perspective projection. The projection coordinates are given by

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \frac{f}{s} \begin{bmatrix} p \\ q \end{bmatrix} \quad (2)$$

where $f$ is the distance of the projection plane of the image to the camera lens, and $[u_0\ v_0]^{\mathrm{T}}$ are the 2-D coordinates of the principal point. Expanding this equation and rewriting it:

$$\frac{x - x_0}{z - z_0} = \frac{r_{11}(u - u_0) + r_{21}(v - v_0) + r_{31} f}{r_{13}(u - u_0) + r_{23}(v - v_0) + r_{33} f} \quad (3)$$

$$\frac{y - y_0}{z - z_0} = \frac{r_{12}(u - u_0) + r_{22}(v - v_0) + r_{32} f}{r_{13}(u - u_0) + r_{23}(v - v_0) + r_{33} f} \quad (4)$$

These equations show that the relation between the 3-D and 2-D coordinates is a function of $u_0$, $v_0$, $x$, $y$, $z$, $\omega$, $\phi$ and $\kappa$, which can be determined through a non-linear least-squares technique, as detailed below.

### A. Least-Squares Estimation

The non-linear least-squares problem is solved by assuming an approximated initial solution, and then iteratively adjusting the partial solutions until a given stopping criterion be achieved. It uses the Newton method, where at each iteration the following equation system is solved:

$$\boldsymbol{\Delta\beta} = [\mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-\mathbf{1}}\mathbf{G}^{\mathrm{T}}\boldsymbol{\varepsilon} \quad (5)$$

where $\varepsilon$ is the vector whose norm represents the objective function to be evaluated, and $\mathbf{G}$ is the Jacobian matrix. The vector $\boldsymbol{\Delta\beta}$ indicates the update to the current solution at $\ell$-th iteration, that is:

$$\boldsymbol{\beta}^{\ell+1} = \boldsymbol{\beta}^{\ell} + \boldsymbol{\Delta\beta}^{\ell} \quad (6)$$

This method usually converges rapidly to a solution, if the initial condition is sufficiently near to the unknown final solution; however the problem can converge to a local minimum instead of to the global minimum.

### B. Exterior and Interior Orientations

In the estimation of the parameters of exterior orientation the $N$ 3-D points having known positions $[x_n\ y_n\ z_n]^{\mathrm{T}}$, $n = 1, \ldots, N$ in the object reference frame are used to obtain the unknown rotation and translation that put the camera reference frame in the world reference frame [5].

The corresponding 3-D point in the camera reference frame and the corresponding 2-D point in the projection plane can be represented as:

$$\begin{bmatrix} p_n \\ q_n \\ s_n \end{bmatrix} = \mathbf{R}(\omega, \phi, \kappa) \begin{bmatrix} x_n - x_0 \\ y_n - y_0 \\ z_n - z_0 \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \frac{f}{s_n} \begin{bmatrix} p_n \\ q_n \end{bmatrix} \quad (8)$$

This problem can be set as a non-linear least-squares problem, that can be solved by assuming an approximated initial solution with the purpose of linearizing the problem. Based on the results of the previous subsection, the linear equation to be solved at each iteration is given by:

$$\boldsymbol{\Delta\beta}^{\ell} = [(\mathbf{G}^{\ell})^{\mathrm{T}}\mathbf{G}^{\ell}]^{-\mathbf{1}}(\mathbf{G}^{\ell})^{\mathrm{T}}(\boldsymbol{\gamma}^{*} - \boldsymbol{\gamma}^{\ell}) \quad (9)$$

where $\gamma^*$ contains the image points, and $\gamma^\ell$ is its estimative at each iteration, that is:

$$\gamma^* = [u_1, v_1, \ldots, u_N, v_N]^{\mathrm{T}} \quad (10)$$

$$\gamma^\ell = [u_1^\ell, v_1^\ell, \ldots, u_N^\ell, v_N^\ell]^{\mathrm{T}} \quad (11)$$

The Jacobian matrix $\mathbf{G}^\ell$ depends upon the 3-D and the 2-D points (see [5]). Eq. (6) is then used for updating the vector $\boldsymbol{\beta} = [x_0\ y_0\ z_0\ \omega\ \phi\ \kappa]^{\mathrm{T}}$ containing the exterior orientation parameters.

The interior orientation is specified by the camera constant $f$ (related to the focal distance), by the coordinates of the principal point $[u_0\ v_0]^{\mathrm{T}}$ (that is the intersection of the optical axis with the image plane), and by the distortion characteristics of the lens used.

### C. Initial Estimation

In order to guarantee a rapid convergence in the exterior orientation computation, an approximated initial solution to the non-linear least-squares procedure is needed in order to compute the exterior orientation parameters. This initial solution is obtained by using a linear method from the projective geometry theory. This method is not robust, being very sensitive to noise, but provides a good initial estimation of the exterior and interior parameters (also called extrinsic and intrinsic parameters).

From [4], it is known that a point $\mathbf{M}$ in the 3-D space is related to its projection $\mathbf{m}$ in the 2-D plane (the image plane) through a linear relation given by

$$\mathbf{m} = \mathbf{PM} \quad (12)$$

where $\mathbf{m} = [\mathbf{U}\ \mathbf{V}\ \mathbf{S}]^{\mathrm{T}}$ and $\mathbf{M} = [\mathbf{X}\ \mathbf{Y}\ \mathbf{Z}\ \mathbf{T}]^{\mathrm{T}}$. This equation is projective, i.e., defined up to a scale factor (given by $S$ and $T$). The matrix $\mathbf{P}$ is the so-called perspective projection matrix. It contains implicitly all the extrinsic and intrinsic parameters. The general form of $\mathbf{P}$ is given below:

$$\mathbf{P} = \begin{bmatrix} \alpha_u \mathbf{r_1} + \mathbf{u_0 r_3} & \alpha_u t_x + u_0 t_z \\ \alpha_v \mathbf{r_2} + \mathbf{v_0 r_3} & \alpha_v t_y + v_0 t_z \\ \mathbf{r_3} & t_z \end{bmatrix} \quad (13)$$

The six extrinsic parameters (three translation parameters and three rotation angles) are obtained from the translation vector $\mathbf{t} = [\mathbf{t_x}\ \mathbf{t_y}\ \mathbf{t_z}]^{\mathrm{T}}$ and from the rotation matrix $\mathbf{R} = [\mathbf{r_1}\ \mathbf{r_2}\ \mathbf{r_3}]^{\mathrm{T}}$. In this work these parameters are used as initial estimation to the least-squares procedure in the computation of the exterior orientation. The intrinsic parameters are given by $\alpha_u$, $\alpha_v$ (related to the focal distance) and $u_0$, $v_0$ (corresponding to the coordinates of the principal point). These parameters are also used in the exterior orientation computation. In order to simplify the reconstruction, these intrinsic parameters are directly used as the parameters of interior orientation, being discarded the effect of lens distortion (see previous section).

### D. A Robust Least-Squares Implementation

As seen before, the exterior orientation computation uses a non-linear least-squares procedure that begins from an approximated initial solution. If a good initial estimative is given, the convergence of the algorithm is rapidly obtained, but in most cases is is not possible to guarantee convergence. The least-squares algorithm is very sensitive to noise in the input points, and sometimes it does not converge to the global minimum, but to a local minimum. Also, the algorithm frequently fails as the number of input points increases. In order to obtain a robust method, some changes, referred next, were done in the least-squares procedure.

In Eq. (9) the estimation of $\boldsymbol{\Delta\beta}$ requires the inversion of the matrix $\mathbf{G}^{\mathrm{T}}\mathbf{G}$. It was observed that this matrix frequently is ill-conditioned, that can lead to singularity. In order to eliminate this problem, the application of the Levenberg-Marquardt approach [7], [8], [9] was used, which consists of the use of a factor $\mu$ to guarantee an acceptable condition number $\mathbf{G}^{\mathrm{T}}\mathbf{G}$. Its value is related to the magnitude of this matrix at each iteration, depending on the error $\boldsymbol{\varepsilon}$ (if the error is reduced, the value of $\mu$ is reduced; otherwise, it is increased until the error decreases).

Then, Eq. (9) is modified to the form:

$$\boldsymbol{\Delta\beta}^\ell = [(\mathbf{G}^\ell)^{\mathrm{T}}\mathbf{G}^\ell + \mu_\ell \mathbf{I}]^{-1}(\mathbf{G}^\ell)^{\mathrm{T}}(\gamma^* - \gamma^\ell) \quad (14)$$

Another important change is in Eq. (6), that represents the update of the parameter vector $\boldsymbol{\beta}$ at each iteration. It was observed that in most cases the step size of each update (related to the rate of convergence) is not adequate. Large steps (i.e., large values of $\boldsymbol{\Delta\beta}$) may lead to local minima or even to divergence. The step size needs to be reduced or increased at each iteration, depending on the direction of the gradient of the evaluation function, and can be controlled by introducing a factor $\alpha_\ell$ in Eq. (6), such that

$$\boldsymbol{\beta}^{\ell+1} = \boldsymbol{\beta}^\ell + \alpha_\ell \boldsymbol{\Delta\beta}^\ell \quad (15)$$

This factor can be made constant for all iterations, although better results are obtained by using a variable step size, modified at each iteration by using the Armijo rules [8], [10]:

1. $F^{\ell+1} \leq F^\ell + \rho_1 \alpha_\ell (\mathbf{g}^\ell)^{\mathrm{T}} \boldsymbol{\Delta\beta}^\ell$, for some $0 < \rho_1 \leq 1$.
2. $(\mathbf{g}^{\ell+1})^{\mathrm{T}} \boldsymbol{\Delta\beta}^\ell \leq \boldsymbol{\rho_2} (\mathbf{g}^\ell)^{\mathrm{T}} \boldsymbol{\Delta\beta}^\ell$, for some $\rho_1 < \rho_2 \leq 1$.

$\mathbf{g} = \mathbf{2G}^{\mathrm{T}}\boldsymbol{\varepsilon}$ is the gradient vector of $F$. The value of $\alpha_\ell$ is chosen accordingly to $\rho_1 \alpha_\ell \leq \alpha_{\ell+1} \leq \rho_2 \alpha_\ell$.

### E. Stereo Triangulation

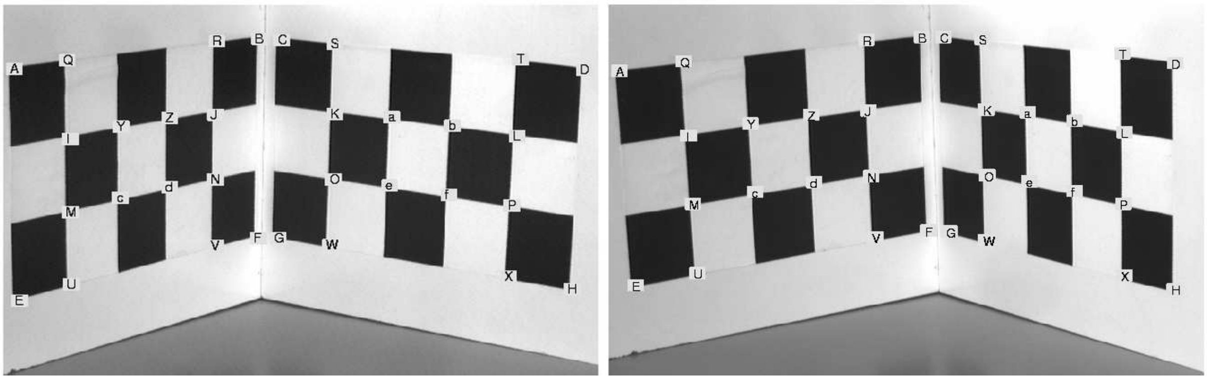
A stereo triangulation procedure is used to obtain the estimation of the 3-D coordinates of a point $[x\ y\ z]^{\mathrm{T}}$, given the 2-D coordinates of its projection on each image of a stereo pair ($[u_L\ v_L]^{\mathrm{T}}$ and $[u_R\ v_R]^{\mathrm{T}}$). The exterior and interior parameters computed for each image are: $[x_L\ y_L\ z_L]^{\mathrm{T}}$, $[x_R\ y_R\ z_R]^{\mathrm{T}}$, $\mathbf{R_L}$ and $\mathbf{R_R}$. The stereo triangulation involves the minimization of the squared difference where $\lambda_L$ and $\lambda_R$ are the parameters to be found in the mentioned minimization:

$$\epsilon^2 = \left\| \begin{bmatrix} x_L \\ y_L \\ z_L \end{bmatrix} + \lambda_L \mathbf{R_L} \begin{bmatrix} u_L \\ v_L \\ f_L \end{bmatrix} - \begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix} - \lambda_{\mathbf{R}} \mathbf{R_R} \begin{bmatrix} u_R \\ v_R \\ f_R \end{bmatrix} \right\|^2 \quad (16)$$

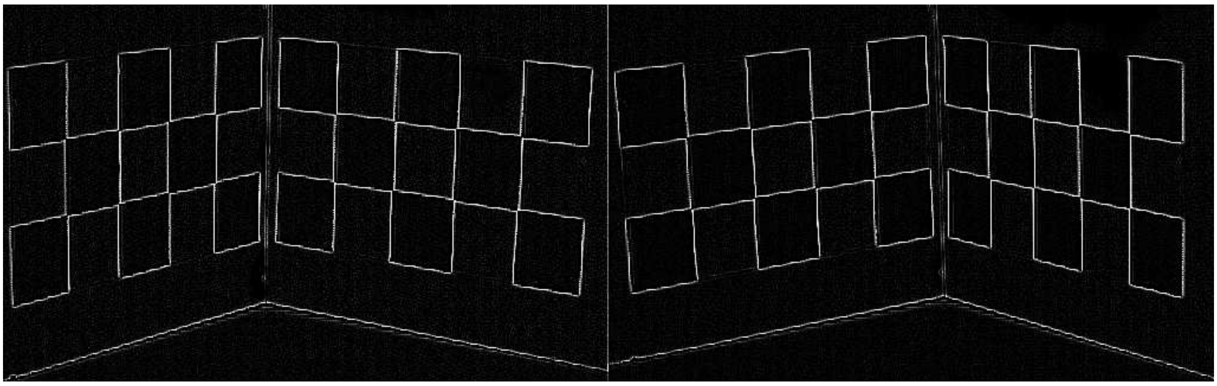
The 3-D coordinates $[x\ y\ z]^{\mathrm{T}}$ are then estimated by using one of the following equations:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_L \\ y_L \\ z_L \end{bmatrix} + \lambda_L \mathbf{R_L} \begin{bmatrix} u_L \\ v_L \\ f_L \end{bmatrix} \quad (17)$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix} + \lambda_R \mathbf{R_R} \begin{bmatrix} u_R \\ v_R \\ f_R \end{bmatrix} \quad (18)$$



(a) original stereo image pair and points used



(b) stereo image pair after edge detection

Fig. 4. Stereo image pair used in the experiments.

## V. Experimental Results

This section illustrates an application of the proposed stereo vision system. Experiments with a stereo image pair have been performed, as shown next. The stereo pair of images used, as well as the result of the edge detection used in the simulation algorithm applied to it, are shown in Fig. 4.

The results of several measurements involving edges of the images are shown in Table 1, where the real and estimated (calculated) dimensions are compared compared, as well the errors (absolute and relative) for each measurement. The points used in each image are labeled in Fig. 4 by capital and small letters. The length of each edge has been estimated from the 3-D coordinates of the vertices that delimit it. The errors are less than 3%, which are considered satisfactory for the application in view.

## VI. Conclusion

This paper presented a stereo vision system designed for activities of inspection and 3-D reconstruction in remote sites. The system was divided into three stages: stereo image acquisition, preprocessing and reconstruction. The vision system allowed the visualization of images in real time, with notion of depth, through a friendly operation interface. The use of edge detection algorithms allowed a better definition in obtaining the points for the computations.

The system described here has been developed for applications in submarine activities in deep waters, having characteristics of low cost, reliability and portability, allowing remote operation in real time with telepresence sensation, which extends its use in some situations where the automation of tasks is necessary, mainly in hostile environments for the involved technicians. It is also suitable for serving as an artificial vision system for autonomous vehicle guidance.

TABLE I
Experimental Results.

| EDGE | REAL DIMENSION | ESTIMATED DIMENSION | ABSOLUTE ERROR | RELATIVE ERROR |
|---|---|---|---|---|
| $\overline{AB}$ | 0.250000 | 0.248476 | -0.001524 | 0.61% |
| $\overline{AE}$ | 0.150000 | 0.147678 | -0.002322 | 1.55% |
| $\overline{IJ}$ | 0.150000 | 0.149961 | -0.000039 | 0.03% |
| $\overline{IM}$ | 0.050000 | 0.048504 | -0.001496 | 2.99% |
| $\overline{QR}$ | 0.150000 | 0.148823 | -0.001177 | 0.78% |
| $\overline{QU}$ | 0.150000 | 0.147848 | -0.002152 | 1.43% |
| $\overline{ab}$ | 0.050000 | 0.049485 | -0.000515 | 1.03% |
| $\overline{Yc}$ | 0.050000 | 0.049489 | -0.000511 | 1.02% |
| $\overline{AF}$ | 0.291548 | 0.288329 | -0.003218 | 1.10% |
| $\overline{AH}$ | 0.397115 | 0.393787 | -0.003327 | 0.84% |
| $\overline{IN}$ | 0.158114 | 0.156998 | -0.001116 | 0.71% |
| $\overline{IP}$ | 0.301164 | 0.300552 | -0.000612 | 0.20% |
| $\overline{QV}$ | 0.212132 | 0.210812 | -0.001320 | 0.62% |
| $\overline{QX}$ | 0.332716 | 0.332996 | 0.000280 | 0.08% |
| $\overline{Yd}$ | 0.070711 | 0.069748 | -0.000962 | 1.36% |
| $\overline{Yf}$ | 0.231733 | 0.231946 | 0.000213 | 0.09% |

## References


[1] Lenildo C. Silva, Antonio Petraglia, and Mariane R. Petraglia, "Stereo Vision System for Live Submarine Inspection of Oil Pipelines and Equipments in Deep Sea," in *Proceedings of the 1998 IEEE Int. Conf. on Intelligent Vehicles – IV'98*, Stuttgart, Germany, October 28–30 1998, pp. 593–598.

[2] A. C. Jacques, J. L. P. Borges, P. C. F. Henrique, and A. G. Sperandio, "Sistema de medição tridimensional para aplicações submarinas - sistema TV3D sem cintilação (TV3D system flickerless)," Tech. Rep., Cenpes, Petrobr´as, Rio de Janeiro, RJ, Brasil, 1994.

[3] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall Inc., 1989.

[4] Olivier Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, The MIT Press, 1993.

[5] Robert M. Haralick and Linda G. Shapiro, *Computer and Robot Vision*, vol. 2, Addison-Wesley Publishing Co., 1993.

[6] Lenildo C. Silva, "Sistema de Visão Tridimensional para Inspeção e Dimensionamento," M.S. thesis, PEE/COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1998.

[7] P. R. Adby and M. A. H. Dempster, *Iintroduction to Optimization Methods*, Chapman and Hall, 1982.

[8] Alan Jennings and J.J. McKeown, *Matrix Computation*, John Wiley & Sons, 2 edition, 1993.

[9] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[10] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, 1995.