STEREO VISION SYSTEM FOR REMOTE MONITORING AND 3-D RECONSTRUCTION

Lenildo C. Silva, Antonio Petraglia, Mariane R. Petraglia

PEE/COPPE-EE, Federal University of Rio de Janeiro
CP 68504, CEP 21945-970, Rio de Janeiro, RJ, Brazil
lenildo@pads.ufrj.br, petra@pads.ufrj.br, mariane@pads.ufrj.br

ABSTRACT

A three-dimensional vision system for inspection activities of installations and equipment by remotely operated vehicles is presented. A real-time vision system is used for the acquisition of stereo pairs of images which, after preprocessing, are submitted to a reconstruction procedure in order to obtain 3-D coordinates and to perform dimensioning of objects. A minimization procedure applied to a robust least-squares implementation of the camera system orientation is described. Experimental results are shown to verify the effectiveness of the proposed techniques.

1. 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 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 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. The three-dimensional vision system can be divided into three stages, as illustrated in Fig. 1, which are described in the following sections.

Figure 1: Block diagram of the real-time stereo system.

2. IMAGE ACQUISITION AND PREPROCESSING

The system is based on a personal computer, and uses two CCD cameras for the stereo image acquisition, making it possible the visualization in real time of the scenes to be investigated [1]. Its architecture is presented in Fig. 2.

The CCD cameras have been assembled on a fixed basis, lined up horizontally, with a baseline of about 5.5 centimeters. Although for reconstruction purposes the baseline should be chosen as large as possible to avoid narrow triangulation, and consequently inaccurate 3-D coordinate estimation, a relatively short value was applied here because it corresponds approximately to the distance between the eyes of an adult human being, and gives the user the sensation of real stereo vision [2]. The digitizer cards are frame grabbers that convert the NTSC video signals coming from the cameras so that they can be shown on a SVGA monitor. The output of both cards are connected to an auxiliary card, whose function is to perform the switching between the left and right images alternatively, at a rate of 120 Hz, to assure that each image is shown for a human eye in 1/60 seconds, thus eliminating blinking effect that would occur in showing the two images alternatively. The stereoscopic imaging device is composed by an LCD screen and passive eyeglasses, that allows the visualization of 3-D scenes from the composition of the 2-D images (left and right), providing the notion of depth.

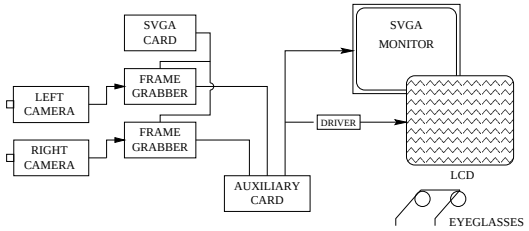

Figure 2: Architecture of the stereo acquisition system.

The preprocessing stage includes techniques of image processing, such as histogram equalization and filtering, to the acquired images. Among the techniques are low-pass filtering for noise smoothing and interpolation and bandpass filtering for the enhancement of edges. Also used is edge detection to allow better characterization of the boundaries of each image, which is particularly useful here for the determination of the 2-D coordinates. A gradient operator is applied to find the edges, followed by a comparison with a threshold to generate a binary edge map [1].

3. RECONSTRUCTION

After the acquisition and preprocessing stages, the stereo images are submitted to the reconstruction stage, where 3-D coordinates are estimated, with the purpose of performing the dimensioning between points of interest in the scene.

The reconstruction procedure comprises photogrammetric methods. Analytic photogrammetry includes a set of techniques by which, from measurements of one or more 2-D perspective projections of a 3-D object, one can make inferences about the 3-D position and orientation of the observed 3-D object in a world reference frame [1, 3, 4].

The complete specification of the orientation of a camera is given by the exterior orientation and by 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 and world reference frames is given by a translation and a rotation. A point $\mathbf{x} = [x\ y\ z]^{\mathrm{T}}$ in the world reference frame is expressed with respect to the camera reference frame by a translation given by the vector $\mathbf{t} = [x_0\ y_0\ z_0]^{\mathrm{T}}$. 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 may differ from the directions of the world reference frame. The rotation by which the world reference frame is positioned with respect to the camera reference frame is a sequence of three rotations, expressed by a rotation matrix $\mathbf{R}(\omega, \phi, \kappa)$.

Thus, a point $[x\ y\ z]^{\mathrm{T}}$ of the world reference frame is represented by the point $[p\ q\ s]^{\mathrm{T}}$ of the camera reference frame, where

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

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

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

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. From (1) and (2) we obtain

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

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

These equations show that the relation between the 3-D and 2-D coordinates is a function of f, u_0, v_0, x_0, y_0, z_0, ω, ϕ and κ, which can be determined through a non-linear least-squares technique, as detailed below.

3.1. Exterior and Interior Orientations Computation

In the estimation of the parameters of the exterior orientation, N 3-D points $[x_n\ y_n\ z_n]^{\mathrm{T}}$ having known positions 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 [3]. The corresponding 3-D points in the camera reference frame and their 2-D projections in the image plane can be determined by (1) and (2).

This problem can be set as a non-linear least-squares approach, which can be solved by first assuming an approximate initial solution (as discussed in Sec. 3.1.1) around which a linear model is produced, and then iteratively adjusting the partial solutions until a given stopping criterium is achieved. It uses the Newton method, so that at the ℓ-th iteration the following equation system is solved:

$$\mathbf{G}^{\ell} \boldsymbol{\Delta\beta}^{\ell} = \boldsymbol{\varepsilon}^{\ell} \tag{5}$$

where $F = \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\varepsilon}$ is the objective function to be evaluated, and $\mathbf{G}$ is the corresponding Jacobian matrix. The vector $\boldsymbol{\Delta\beta}^{\ell} = [\Delta x_0^{\ell}\ \ \Delta y_0^{\ell}\ \ \Delta z_0^{\ell}\ \ \Delta\omega^{\ell}\ \Delta\phi^{\ell}\ \Delta\kappa^{\ell}]^{\mathrm{T}}$ indicates the update to the current solution:

$$\boldsymbol{\Delta\beta}^{\ell} = [(\mathbf{G}^{\ell})^{\mathrm{T}} \mathbf{G}^{\ell}]^{-1} (\mathbf{G}^{\ell})^{\mathrm{T}} (\boldsymbol{\gamma}^{*} - \boldsymbol{\gamma}^{\ell}) \tag{6}$$

The Jacobian matrix $\mathbf{G}$ depends upon the 3-D and the 2-D points [3], and $\boldsymbol{\varepsilon} = \boldsymbol{\gamma}^{*} - \boldsymbol{\gamma}^{\ell}$. The vector $\boldsymbol{\gamma}^{*}$ contains the image points, while $\boldsymbol{\gamma}^{\ell}$ is its estimate at ℓ-th iteration. Expanding the vectors:

$$\boldsymbol{\gamma}^{*} = [u_1\ v_1\ \cdots\ u_N\ v_N]^{\mathrm{T}} \tag{7}$$

$$\boldsymbol{\gamma}^{\ell} = [u_1^{\ell}\ v_1^{\ell}\ \cdots\ u_N^{\ell}\ v_N^{\ell}]^{\mathrm{T}} \tag{8}$$

The vector $\boldsymbol{\beta} = [x_0^{\ell}\ \ y_0^{\ell}\ \ z_0^{\ell}\ \ \omega^{\ell}\ \phi^{\ell}\ \kappa^{\ell}]^{\mathrm{T}}$ containing the parameters is updated at each iteration by using the equation

$$\boldsymbol{\beta}^{\ell+1} = \boldsymbol{\beta}^{\ell} + \boldsymbol{\Delta\beta}^{\ell} \tag{9}$$

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

3.1.1. Initial Estimation

In order to guarantee a fast convergence of the exterior orientation computation, an approximate initial solution to the non-linear least-squares procedure is needed, as mentioned previously. This initial solution is obtained by using a linear method from the projective geometry theory. Although sensitive to measurement noise, this method provides good initial estimates for the extrinsic and intrinsic parameters.

A point $\mathbf{M}$ in the 3-D space is related to its projection $\mathbf{m}$ in the image plane through a linear relation given by [5]

$$\mathbf{m} = \mathbf{P}\mathbf{M} \tag{10}$$

where $\mathbf{m} = [U, V, S]^{\mathrm{T}}$ and $\mathbf{M} = [X, Y, Z, T]^{\mathrm{T}}$. This equation is projective, i.e., defined up to a scale factor (S and T in the above vectors). The matrix $\mathbf{P}$ is the so-called perspective projection matrix, and contains implicitly all the exterior orientation and interior orientation parameters, referred

here as extrinsic and intrinsic parameters, respectively. The general form of $\mathbf{P}$ is

$$\mathbf{P} = \begin{bmatrix} \alpha_u \mathbf{r}_1 + u_0 \mathbf{r}_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v \mathbf{r}_2 + v_0 \mathbf{r}_3 & \alpha_v t_y + v_0 t_z \\ \mathbf{r}_3 & t_z \end{bmatrix} \quad (11)$$

The six extrinsic parameters are obtained from the translation vector $\mathbf{t} = [t_x\ t_y\ t_z]^{\mathrm{T}}$ and from the rotation matrix $\mathbf{R} = [\mathbf{r}_1^{\mathrm{T}}\ \mathbf{r}_2^{\mathrm{T}}\ \mathbf{r}_3^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ are row-vectors of $\mathbf{R}$. In this work these parameters are used as initial estimation in the least-squares procedure for the computation of the exterior orientation. The intrinsic parameters are given by α_u, α_v (related to the focal distance), u_0, v_0 (the coordinates of the principal point). These parameters are also used in the exterior orientation computation. In order to simplify the reconstruction procedure, these intrinsic parameters are directly used as the parameters of interior orientation.

The parameters related to lens distortion can be non-linearly estimated using the model

$$\begin{cases} u' = u + \delta_u(u, v) \\ v' = v + \delta_v(u, v) \end{cases} \quad (12)$$

where $[u\ v]^{\mathrm{T}}$ are the distortion-free image coordinates, and $[u'\ v']^{\mathrm{T}}$ are the corresponding coordinates with distortion. $\delta_u(u, v)$ and $\delta_v(u, v)$ are the distortion amounts considering both geometric and tangential distortion effects [6]:

$$\begin{cases} \delta_u = (g_1 + g_3)u^2 + g_4 uv + g_1 v^2 + k_1 u(u^2 + v^2) \\ \delta_v = g_2 u^2 + g_3 uv + (g_2 + g_4)v^2 + k_1 v(u^2 + v^2) \end{cases} \quad (13)$$

3.1.2. A Robust Least-Squares Implementation

If a good initial estimate is produced, the convergence of the algorithm is rapidly achieved, but in most cases it is not possible to guarantee convergence. The least-squares algorithm is 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, described next, were included in the least-squares procedure.

In Eq. (6) the estimation of $\boldsymbol{\Delta\beta}$ requires the inversion of the matrix $\mathbf{G}^{\mathrm{T}}\mathbf{G}$. It was observed that this matrix is frequently ill-conditioned, leading to singularity. To eliminate this problem, the Levenberg-Marquardt approach [7] was applied, by adding a factor μ to the diagonal elements of $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ to guarantee an acceptable condition number for this matrix. Its value is related to the magnitude of this matrix and to the objective function $F = \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}$ at each iteration. If $F^{\ell+1} \geq F^{\ell}$ the value of μ_ℓ is increased until $F^{\ell+1} < F^{\ell}$ is reached. Accordingly, Eq. (6) is modified as:

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

Another important change is in Eq. (9), which 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 α_ℓ in Eq. (9):

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

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 [7]

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

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

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

3.2. 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 its 2-D projections $[u_L\ v_L]^{\mathrm{T}}$ and $[u_R\ v_R]^{\mathrm{T}}$ on each image of a stereo pair. 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 λ_L and λ_R are the parameters to be found in the mentioned minimization, i.e.:

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

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

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

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

3.3. Stereo Matching Procedure

The image points used to estimate the exterior and interior orientations, as well as the image points needed in the stereo triangulation procedure, are selected semi-automatically in the stereo images by applying a stereo matching procedure: the operator selects the required points in the left image, and automatically their corresponding points are identified in the right image.

The method adopted is a correlation-based method, that finds conjugate patches of two images having similar brightness patterns. Given a patch in an image, typically a single rectangular window, the correlation with each patch along the corresponding epipolar line on the other image is computed, and the point having the "best" correlation value is accepted as the corresponding point [8]. The correlation

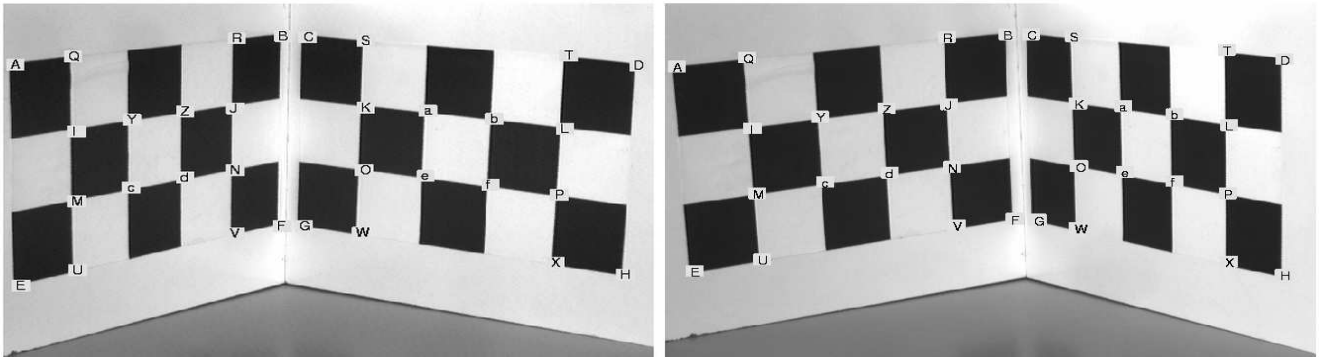

Figure 3: Stereo image pair used in the experiments.

Table 1: Experimental Results.

EDGE	REAL DIMENSION	ESTIMATED DIMENSION	ABSOLUTE ERROR	RELATIVE ERROR
$\overline{AB}$	0.250000	0.248476	-0.001524	0.61%
$\overline{CG}$	0.150000	0.148061	-0.001939	1.29%
$\overline{IJ}$	0.150000	0.149961	-0.000039	0.03%
$\overline{IM}$	0.050000	0.048504	-0.001496	2.99%
$\overline{SW}$	0.150000	0.147683	-0.002317	1.54%
$\overline{ab}$	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%
$\overline{Yf}$	0.231733	0.231946	0.000213	0.09%

is computed considering the sum of squared-differences, given by:

$$C(u)=\sqrt{\sum_i \sum_j \big(I_L(u_L+i, v_L+j)-I_R(u+i, v_R+j)\big)^2} \quad (19)$$

where i and j are, respectively, the column and row positions in the patch. Since the camera planes are parallel to each other, all the epipolar lines are also parallel to each other along the x-axis, yielding $v_L = v_R$. Therefore, the scanning performed to match the point $[u_L\ v_L]^{\mathrm{T}}$ in the right image is made along the epipolar line located at $v = v_R$.

4. EXPERIMENTAL RESULTS

This section presents an illustrative application of the proposed stereo vision system. The stereo pair of images used in the experiments are shown in Fig. 3. The results of several measurements involving edges of the images are shown in Table 1, where the real and estimated (calculated) dimensions are compared. The points chosen in each image are labeled in Fig. 3 by capital letters. The dimension 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.

5. CONCLUSIONS

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.

Although the system described here has been originally developed for applications in submarine activities in deep waters, its characteristics of low cost, reliability and portability, allow remote operation in real time with telepresence sensation in many situations where the automation of tasks is necessary, mainly in hostile environments for the user. It is also potentially suitable for serving as an artificial vision system for autonomous vehicle guidance.

6. REFERENCES

[1] L. C. Silva, A. Petraglia, and M. R. Petraglia, "Stereo Vision System for Live Submarine Inspection of Oil Pipelines and Equipments in Deep Sea," in *Proc. IEEE Int. Conf. on Intelligent Vehicles*, Stuttgart, Germany, Oct. 1998, pp. 593–598.

[2] M. Okutomi and T. Kanade, "A Multiple Baseline Stereo," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 15, no. 4, pp. 353–363, Apr. 1993.

[3] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 2, Addison-Wesley Publishing Co., 1993.

[4] C.-P. Lu, E. Mjolsness, and G. D. Hager, "Online Computation of Exterior Orientation with Application to Hand-Eye Calibration," *Mathl. Comput. Modeling*, vol. 24, no. 5/6, pp. 121–143, 1996.

[5] O. Faugeras, "Stratification of 3-D Vision: Projective, Affine and Metric Representations," *J. of the Opt. Soc. of America*, vol. 12, no. 3, pp. 465–484, Mar. 1995.

[6] J. Weng, P. Cohen, and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992.

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

[8] A. Redert, E. Hendriks, and J. Biemond, "Correspondence Estimation in Image Pairs," *IEEE Signal Processing Magazine*, vol. 16, no. 5, pp. 29–46, 1999.