Proceedings of the XI DINAME, 28th February-4th March, 2005 - Ouro Preto - MG - Brazil
Edited by D.A. Rade and V. Steffen Jr.

3-D IMAGE SIGNAL PROCESSING FOR AUTOMATED OPERATIONS USING A RANGE CUBE

Paulo Roberto Gardel Kurka
DPM/FEM/UNICAMP
Caixa Postal 6122, Campinas – SP
CEP 13084-971
kurka@fem.unicamp.br

Marcelo Rudek
Lboratório de Automação e Sistemas
Engenharia de Produção e Mecatrônica
Pontifícia Universidade Católica do Paraná
marcelo.rudek@pucpr.br

Luiz Ferreira Menezes Jr.
CPG/FEM/UNICAMP
Caixa Postal 6122, Campinas – SP
CEP 13084-971
luiz@eletronz.com

Abstract: *The work shows the progress of a procedure to determine the position and volumetric information of an object, in the space of a work cell. The determination is done based in area mapping through a virtual cube. This text The work which has applications in robotics, and manufacturing is a cooperation effort involving UNICAMP and PUCPR. Aspects of camera calibration geometry and the practical processing of signals are addressed in the paper.*
Keywords: *Image Processing, 3-D Reconstruction, Signal Processing.*

1. Introduction

Modern production processes require computational and robotic systems to execute many complex tasks in an automated way. In this context, computer vision systems and image processing are used in production cells, to aid tasks such as robot orientation and movement, positioning of objects, quality control, and others.

In a robotic environment based on vision systems, the configuration of the work cell involves the positioning of cameras, in order to provide image maps of objects from the real 3-D world. The process of relating image maps with 3-D objects is called camera calibration, and is based on the representation of the scenario points in each camera. There are basically two ways of using automatic vision: via fixed cameras in the automated work cell or through a robot “hand-mounted” camera. The advantage of using a camera mounted on a robotic arm, is the obtention of a good visual field during the time of fly, but such a configuration requires the recalibration for each position during the movement, as presented by Mota (2002). The fixed configuration, has the advantage of estimating the calibration matrices only once, since there are no changes in the environmental configuration.

The present work uses a fixed cameras model, in order to determine geometric aspects of objects, such as volume and position. It presents the mathematical model of a camera, as well as its calibration process. Vision systems can be set up with one or more cameras. Although it is possible to have a a convenient map of the real 3D word using a single camera and a segmentation object, as described by Ryoo (2004), a more flexible and efficient mapping system is based on stereo vision, using a pair of calibrated cameras. The work also proposes a new way of representing a solid position and dimension, based on referencing cubes.

Estimation of the essential matrix is a preliminary step in obtaining a 3-D model of correspondent objects in a pair of stereoscopic images. The essential matrix can be decomposed, using singular value decomposition (SVD) in order to recover the rotation and translation parameters, within a non-iterative solution Tsai (1984), Fiore (2000). The parametric rotation matrix $\boldsymbol{R}$ and the origin translation vector $\boldsymbol{t}$ are the relative positions of the cameras used for capturing the images. It is possible to calculate depth measurements from a pair of images once their rotation matrix and origin translation vector have been estimated in an adequated way Trucco (1998).

Procedures for estimation of the essential matrix are found in 3D reconstruction applications, where metric accuracy is requested. Different estimation approaches are presented by Salvi (2002), in the context of camera calibration methods. Borghese (2000), describes in a simple manner, a technique based on the previously known information of a small number of matched image points, associated with epipolar transformations Luong (1992). Lourakis (2000) suggests two strategies for indicating matching points in an image pair. The first assumes an arbitrary knowledge of geometric constraints, and the second exploits the projective quantities that remain invariant. Image matching can also be performed using a bilinear relation as presented by Faugueras (2000), or through a Harris and Stephens detector, such as the one used by Oisel (2003). The Harris and Stephens detector is based on a corner detector function, created by Moreavec, as cited in Bas (2002). In Moreavec's theory, candidate-matching points in an image must have a high variance, such as the corners of an image object. Such a detector is a robust method based on autocorrelation measure, as explained in Rockett (2002).

An interesting method for the reconstruction of a 3D model from weakly calibrated images, with previous estimation of the essential matrix, is presented by Oisel (2003). The proposed method for estimating the essential matrix is to be applied in similar situations where a stereoscopic pair of images, or two different images obtained from a moving camera, is known, for a given solid. Such applications are typical of a solid measurement/positioning verification system in an assembly line Rudek (2000), or obstacle avoidance procedures for mobile robots in real environments Becker (2003).

The geometric background to enable the estimation of the essential matrix from camera information and the proposed range cube thechnique is presented in this work, in the sequence that follows. Section 2 presents the epipolar geometry relations that are the basis for 3-D image reconstruction. Section 3 presents the forming of the essential matrix, as well as an algorithm to estimate it from a number of known image pairs. Section 4 describes the range cube aplication and section 5 presents the main conclusions of this work.

2. Image formation

The mathematical representation of an image can be done through a relationship model between the world coordinates of a 3D point and its corresponding coordinates on the image plane. A way to perform this relationship, is to use a "pinhole" model, as in Fusiello (2000). Such a model is presented in figure 1, where a 3D point $\boldsymbol{P}^{w}(\mathbf{X,Y,Z})$, it is projected onto a point $\boldsymbol{p}'(\mathbf{x,y})$ on a 2D image plane.

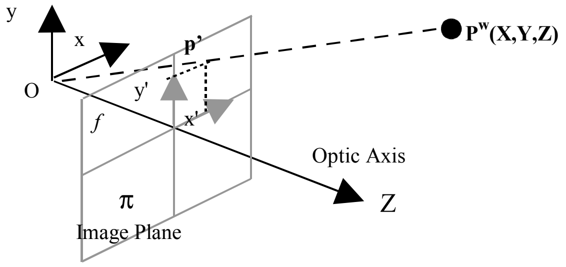


Figure 1. Geometric representation of the 3D point in the image plane.

The model consists of an image plane π and a reference point **O** in the 3D space, called the projection centre. The distance ***f*** between the image plane and **O** is the focal distance. The line trough **O** and perpendicular to π is the optic axis. Point ***p'*** is where the straight line $\overline{P^w O}$, crosses the image plane. The 3D coordinate frame is ***P***= $[X,Y,Z]^T$ and the frame of camera in metric units is ***p'*** = [x',y',f]. Therefore, it can be writen that,

$$\boldsymbol{p'} = \boldsymbol{R}\,\boldsymbol{P} + \boldsymbol{t}. \tag{1}$$

Using QR factorization as in Fusiello (2000) and Oisel (2003), leads to

$$\tilde{\boldsymbol{p}} = [\boldsymbol{R} \mid \boldsymbol{t}] \tag{2}$$

which is the perspective projection matrix in homogeneous coordinates. The camera is then modeled by its rotation (**R**) and translation (***t***) matrices, which represents its extrinsic parameters. ***A*** is the matrix of intrinsic parameters, and contains the internal camera properties.

2.1. Epipolar Geometry

Under 3D reconstruction, it necessary to work with two images and to exploit their disparity for obtaining depth information. However, establishing correspondences of points between two images is a difficult task. The epipolar constraints helps to limit the searching regions for matching points.

The work presented by Yi Ma (2004), considers two images of the same point p, taken from two distinct cameras, with centres in O_1 and O_2, and image planes parallel to the xy plane, as seen on figure 2. The intersections of line $\overline{o_1 o_2}$ with each image plane are called *epipoles* and denoted by e_1 and e_2. Lines ℓ_1 and ℓ_2 are called *epipolar lines*, which are the intersection of the plane $o_1 o_2 p$ with the two image planes.

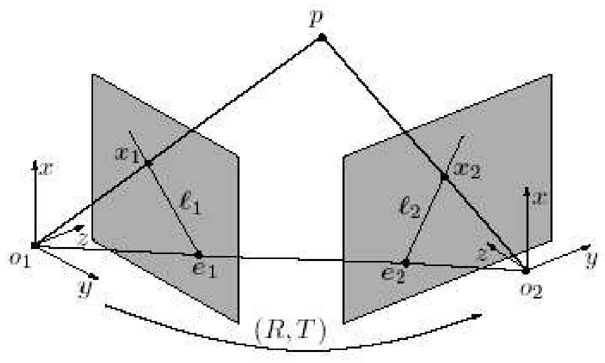


Figure 2 – Geometry of two image projections by YI Ma (2004),

It can be seen, from the epipolar geometry relations, that all points along the line $\overline{o_1\, p}$ have the same projection point x_1 in the image plane of camera 1. Therefore, it can also be said that the single projection point in camera 1 corresponds to the epipolar line ℓ_2 in camera 2.

Points x_1 and x_2 are the projections of point *p* on the image planes of both cameras. In spatial terms, the projection points can be represented by 3-D vectors, $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, with coordinates *x*, *y* and *f*, relative frame systems placed at camera 1 or 2, respectively:

$$\boldsymbol{x_1} = \begin{bmatrix} x_1 \\ y_1 \\ f \end{bmatrix} \quad , \quad \boldsymbol{x_2} = \begin{bmatrix} x_2 \\ y_2 \\ f \end{bmatrix} \tag{3}$$

where *f* is the focal distance of the image plane.

Camera 2 has a rotation matrix ***R***, which contains the orientation of each axis, according to a system of reference coordinates, arbitrarily fixed at camera 1:

$$\boldsymbol{R}=\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{4}$$

where

$$\begin{array}{lll} r_{11}=\cos\phi\cos k & r_{12}=\sin\omega\sin\phi\cos k+\cos\omega\sin k & r_{13}=-\cos\omega\sin\phi\cos k+\sin\omega\sin k \\ r_{21}=-\cos\phi\sin k & r_{22}=-\sin\omega\sin\phi\sin k+\cos\omega\cos k & r_{23}=\cos\omega\sin\phi\sin k+\sin\omega\cos k \\ r_{31}=\sin\phi & r_{23}=-\sin\omega\cos\phi & r_{33}=\cos\omega\cos\phi \end{array}$$

and ω – Rotation about x axis, θ – Rotation about y axis, k – Rotation about z axis.

Translation vector $\boldsymbol{t}$ gives the relative position of the two camera centres. A point p in the 3-D space, has coordinates $\mathbf{X}_1$ and $\mathbf{X}_2$, with respect to relative frame systems placed at camera 1 or 2, respectively. Such coordinates are related by a rigid body transformation in the following way:

$$\boldsymbol{X}_2=\boldsymbol{R}\boldsymbol{X}_1+\boldsymbol{t} \tag{5}$$

The spatial and camera projection coordinates of point p, are related in terms of the depth parameter λ in the following way:

$$\boldsymbol{X}_i=\lambda_1\boldsymbol{x}_i, \qquad i=1,2 \tag{6}$$

It can be seen, from equations (5) and (6) above, that the 3-D coordinates of a point can be established, provided its projections in cameras 1 and 2, as well as the rotation matrix and translation vector, are all known. In practical applications of 3-D image reconstruction, camera projections of spatial points are the known quantities. Relative camera rotation and origin translation vector are not a priori known. The estimation of such quantities is the starting point on 3-D reconstruction techniques.

3. Essential Matrix Estimation

The epipolar constraint can be represented by a matrix, called the *essential matrix,* when the intrinsic parameters of the cameras are known. Otherwise, it is called *fundamental matrix*, as presented by Forsyth (2003).

Equation (5) can be written in terms of projection coordinates and depths:

$$\lambda_2\boldsymbol{x}_2=\boldsymbol{R}\lambda_1\boldsymbol{x}_1+\boldsymbol{t} \tag{7}$$

A translation cross product matrix $\hat{\boldsymbol{T}}$ can be created as:

$$\hat{\boldsymbol{T}}=\begin{bmatrix} 0 & -t_x & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{8}$$

where t_x, t_y and t_z are the coordinates of the translation vector $\boldsymbol{t}$. Matrix $\hat{\boldsymbol{T}}$ has the following properties:

$$\hat{\boldsymbol{T}}\boldsymbol{t}=\boldsymbol{0}, \qquad \hat{\boldsymbol{T}}\boldsymbol{x}=\hat{\boldsymbol{T}}\otimes\boldsymbol{x} \tag{9}$$

Premultiplying both sides of equation (7) by $\hat{\boldsymbol{T}}$ yields:

$$\lambda_2\hat{\boldsymbol{T}}\boldsymbol{x}_2=\hat{\boldsymbol{T}}\boldsymbol{R}\lambda_1\boldsymbol{x}_1 \tag{10}$$

Premultiplying equation by the vector $\boldsymbol{x}_2^T$ gives:

$$\boldsymbol{x}_2^T\boldsymbol{E}\boldsymbol{x}_1^T=\boldsymbol{0} \tag{11}$$

where $\mathbf{E}$ is the essential matrix defined as:

$$\boldsymbol{E}=\hat{\boldsymbol{T}}\boldsymbol{R} \quad . \tag{12}$$

The essential matrix, comprised of the parameters of rotation and translation of the camera reference systems, is the information to be retrieved in a 3-D image reconstruction application. Equation (11) is also known as the *essential constraint* relation.

It is assumed that the essential matrix has the form:

$$\boldsymbol{E}=\begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} . \tag{13}$$

A stacked form of matrix **E** is given as:

$$(\boldsymbol{E}^{\boldsymbol{S}})^{\boldsymbol{T}}=[e_{11} \quad e_{21} \quad e_{31} \quad e_{12} \quad e_{22} \quad e_{32} \quad e_{13} \quad e_{21} \quad e_{33}]$$

Thus, the following homogeneous system of linear equations can be formed:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{19} \\ a_{21} & a_{22} & \cdots & a_{29} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k9} \end{bmatrix} \boldsymbol{E}^{\boldsymbol{S}}=\boldsymbol{0} , \tag{14}$$

where a_{kj} is the j-th element of the Kronecker product, from the pair "*k*" of two projection vectors of the type $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$a_k=[a_{k1} \quad a_{k2} \quad \cdots \quad a_{kj} \quad \cdots \quad a_{k9}]=\boldsymbol{x}_{\mathbf{1}} \otimes \boldsymbol{x}_{\mathbf{2}} . \tag{15}$$

A non trivial solution for the homogeneous system of equations is found, for a number $k \geq 8$ of different pairs of projection points. Furthermore, the *k* pairs of projected points must be originated in a non-coplanar surface of the 3-D space. A solution for $\boldsymbol{E}^S$ if found by computing the eigenvector of the coefficients matrix of the homogeneous system of equations, which corresponds to its smallest eigenvalue. A solution to the essential matrix $\boldsymbol{E}$, still has to be derived from the projection of the stacked vector $\boldsymbol{E}^S$ onto the space of defined essential matrices.

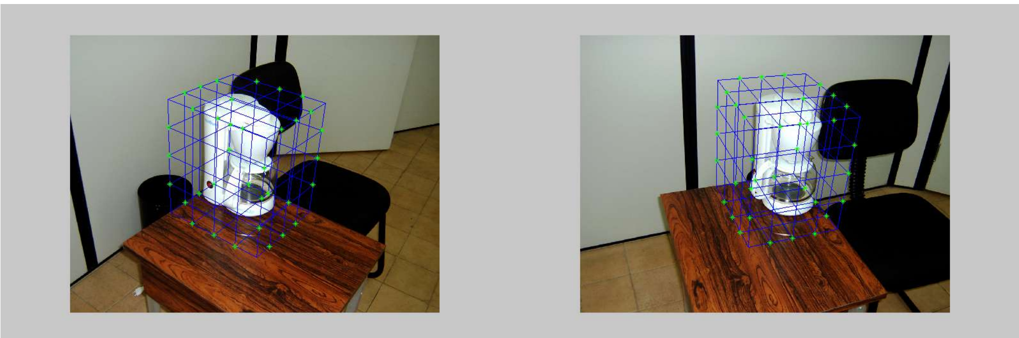
From the estimated essential matrix, it is possible to recover the camera's rotation parameters matrix $\boldsymbol{R}$, and a unit-normalized vector $\boldsymbol{t}$. The module of distance between the cameras origins must be known, in order to perform the estimation of the 3-D position of the projected points.

The procedure described above is formulated in terms of ideal camera parameters and image projections. In practice, although the basic 3-D coordinate estimation procedures remain the same, some pre-conditioning of the projected parameters must be performed in order to compensate for lenses distortion, camera origin offset and pixel deformation.

4. Solid Positioning Determination

Determination of a solid position and dimensions is an important feed-back information in automated processes. The present work proposes a method to obtain volume and positioning information of a solid, using the mapping of a limited space around it. This space is mapped in the form of a virtual cube. A cubic calibration grid is used in the region of interest for measuring the solid positioning and dimension.

An initial grid is obtained by processing the images of a solid calibration cube, and identifying on both images corresponding points for camera calibratoin. Calibration and initial grid points can be those of the cube's visible vertices, mid surfaces and mid edges. After removal of the calibrating cube, its verices and edges are used as references for generating a finer virtual 3-D mesh,which contains the object to be mesaured. Figure 3, presents two views of an object with the superimposed virtual cubic mesh.



Left Right

An automated boundary location algorithm, based on epipolar geometry, is to be used in order to determine which of the virtual cubes are occupying the same positions as the solid, and which are not. An estimation of the solid's volume occupied by elementary cubes, as well as its location on the envolving prism, provides information on the solid's dimension and positioning.

5. Conclusion

The paper displays the progress of studies to estimate the position and dimensions of a solid in a manufacturing cell. The background to image procesing theory is presented, arriving to the important point of estimation of the essential matrix, which is the base for 3-D reconstruction techniques.

Practical use of the 3-D reconstruction from a pair of stereoscopic images is proposed by means of definition of a virtual cubic grid around the solid whose position and dimensions are to be determined. Such a procedure is convenient for the measurement of solids in an automated assembly line, with a pair of fixed cameras, focused on the regin of measuring interest.

The work is under progress and the first practical results are about to be obtained.

6. References

Salvi, J.; Armangué, X.; Batlle, J., 2002. A Comparative Review of Camera Calibrating Methods With Accuracy Evaluation; Pattern Recognition, 35, 1617-1635.

Borghese, N. A.; Cerveri, P., 2000. Calibrating a Video Camera Pair With a Rigid Bar; Pattern Recognition, 33, 81-95.

Oisel, L.; et al, 2003. One Dimensional Dense Disparity Estimation for Three-Dimensional Reconstruction; IEEE Transactions on Image Processing, vol 12, no.9.

Luong, Q.-T.; Faugueras, O. D., 1992. Self-calibration of a camera using multiples images; IEEE.

Lourakis, M. I. A.; et al, 2000. Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing Planes. VIRGO research network of the TMR Programme. February.

Faugueras, O.; Quan, L.; Strum, P., 2000. Self-Calibration of a 1D Projective Camera and Its Aplication to the Self-Calibration of a 2D Projective Camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no.10.

Bas, P.; Chassery, J-M.; Macq, B., 2002. Geometrically Invariant Watermarking Using Feature Points. IEEE Transactions on Image Processing , Vol 11, 9, pp 1014-1028.

Rockett, P. I., 2002. Performance Assessment of Feature Detection Algorithms: A Methodology and Case Study on Corner Detectors. Department of Electronic and Electrical Engineering, University of Sheffield, UK.

Tsai, R.; Huang, T., 1984. Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surface. IEEE Trans. Pattern Anal. Machine Intell., vol 6, pp.13-26.

Fiore, P. D., 2000. Linear Epipolar Algorithm for Multiframe Orientation; IEEE, pp.557-560, 2000;

Trucco, E.; Verri, A., 1998. Introductory Techniques for 3-D Computer Vision; Prentice Hall.

Becker, M; Rodrigues, D. L. 2003 - Detection, Classification, and Tracking of Obstacles in Dynamically Changing Environments using Laser Scanning- Proceedings of the X International Symposium on Dynamic Problems of Mechanics, pp 261-266, ABCM.

Motta, J. M. S. T.; McMaster, R. S., 2002. Experimental Validation of a 3-D Vision-Based Measurement System Applied to Robot Calibration. Journal of the Braz. Soc. Mechanical Sciences, vol. XXIV, p. 234-238, July.

Ryoo, Seung Taek, 2004. Segmentation Based Environment Modeling Using a Sigle Image. International Conference, ICIAR 2004, Proceedings, Part I, p.98-105.

Fusiello, A., 2000. Uncalibrated Euclidian Reconstruction: A Review. Image and Vision Computing, v 18, p.555-563.

Forsyth, D. A; Ponce, J., 2003. Computer Vision. A Modern Approach. Prentice Hall.

Ma, Y., Soatto, S., Kosecká, J., Sastry, S.S., 2004 An Invitation to 3-D Vision From Images to Geometric Models Series: Interdisciplinary Applied Mathematics, Vol. 26, Springer. 526 p.

Rudek, M; 2000 Positioning and Dimensioning of solids in an Assembly Line, Through Image Analysis. Doctorate Project at the Faculty of Mechanical Engineering, UNICAMP.

7. Responsibility notice

The authors are the only responsible for the printed material included in this paper.