3-D MEASUREMENT FROM IMAGES USING A RANGE BOX

Marcelo Rudek
PRODUTRONICA/ PUC PR
Grupo de Eng. De Produção e Mecatrônica
CEP 80215-901
marcelo.rudek@pucpr.br

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

Abstract. *The work shows the progress of a procedure to determine the position and volumetric information of a solid object in the space of a robotic working cell. The proposed method uses a pair of images to obtain a 3-D mapping of the space around the solid. A virtual range box is initially positioned around the object. The box is adjusted around the object's image projection border through perspective translation of its faces. The position and volume of the solid is estimated by reconstruction of the adjusted range box. The proposed technique has applications in robotics and manufacturing, and involves a research cooperation project between UNICAMP and PUCPR. Aspects of camera calibration, 3-D reconstruction and the practical processing of images are addressed in the paper.*

Keywords*:* computer vision, 3-D reconstruction, image processing, range box

1. Introduction

Image acquisition systems are the object of interest in automated industrial applications, for the search of 3-D information that will take part in more reliable and cost efficient process and final product inspection tasks, increasing also the quality of the produced goods. Many theories, techniques, mathematical and computational tools are being developed for recovering of 3-D information from image analysis. The extraction of 3-D metric relation from projection images can be also used in the navigation control of autonomous robots. Among various propositions in this field, the present work displays a convenient way to determine the approximated volume and positioning of a solid, based on a referencing box (Kurka *et al*, 2005).

Vision systems can be set up with one or more cameras. Although it is possible to have a convenient map of the real 3D world using a single image and a segmentation process, as described by Ryoo (2004), a more flexible and efficient mapping system is based on two projected views, using a pair of calibrated cameras. The present work considers the pinhole camera configuration (Trucco and Verri, 1998, Forsyth and Ponce, 2003, Y Ma *et al*, 2004), in order to determine geometric aspects of objects and its position and dimension. It presents the mathematical model of a camera, as well as the process of relating image maps with 3-D objects, known as the calibration process.

Estimation of the essential matrix is a preliminary step in obtaining a 3-D model of correspondent objects in a pair of images within a process known as rigid body transformation (Forsyth and Ponce, 2003). The essential matrix can be obtained, using singular value decomposition (SVD) in order to recover the rotation and translation parameters, within a non-iterative solution as in Tsai (1984) and Fiore (2000). The parametric rotation matrix ***R*** and the origin translation vector ***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.

Procedures for estimation of the essential matrix are found in 3-D 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, an estimation technique based on previously known information of a small number of matched image points, associated with epipolar transformations, also as in Luong (1992). The matching points can be obtained manually or in an automatic fashion. The difficulties of automatic matching have been presented in many works. Lourakis (2000) suggests the assumption of arbitrary geometric constraints, and 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).

These automatic processes are necessary when the cameras have constant movement, changing their positions, and the rotation and translation matrices must be recalculated at each time. In the context of the present work, it is assumed that the environment is not changing and the solid object of interest is found in the area mapped by a pre defined range box. The range box is a referencing figure, located over the mapping area. The range box is modified until its new projections circumscribe a desired region of interest. The adjusted range box vertices are used to obtain the 3-D estimates of position and volume occupied by the object in the area of interest.

The geometric background that describes the essential matrix and the proposed range box technique 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 box application and section 5 presents the main conclusions of this work.

2. Epipolar Geometry Background

The computing of an object's position in space requires mathematics equations that link the 3-D regions with their corresponding image point projection from calibrated camera parameters. Such equations are described in the camera reference frame and the point projections are given in pixel coordinates within the same reference system. Recovering of the 3-D information from a pair of 2-D projection is therefore made from a previous knowledge of the camera's internal characteristics (known as intrinsic parameters) and its position information (called extrinsic camera parameters or "pose"). For this study, the intrinsic parameters are assumed to be known and the extrinsic parameters are obtained from a suitable camera calibration procedure, as the one shown in Kurka *et al* (2005). The cameras pose are represented by their rotation (**R**) and translation (**t**) matrices.

Information for reconstruction of a 3-D map from two projected views is commonly achieved through examination of a disparity function, which yields convenient depth information for each pair of projected points (Oisel, 2003). However, establishing correspondences of points between two images is a difficult task. The epipolar constraints help to limit the searching regions for matching points.

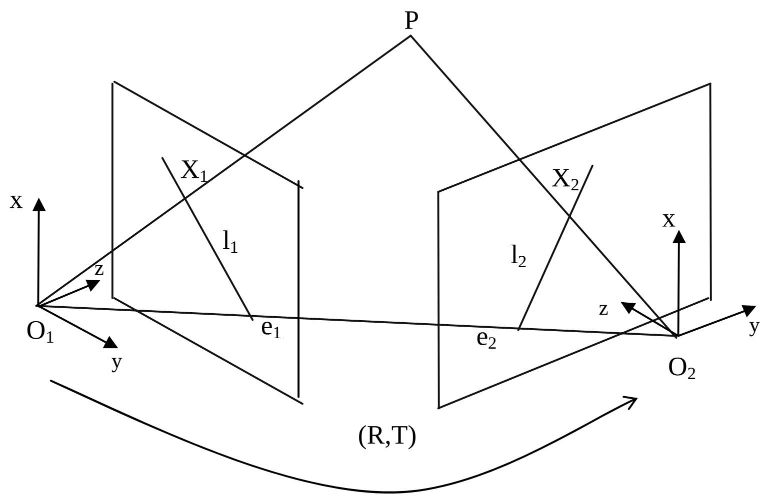


Figure 1. Epipolar arrangement for two views.

In figure 1, $\mathbf{X}_1$ and $\mathbf{X}_2$ are images of point **P,** observed by two cameras with optical centre $\mathbf{O}_1$ e $\mathbf{O}_2$. The lines $\mathbf{O}_1\mathbf{P}$ and $\mathbf{O}_2\mathbf{P}$ form a plane called the epipolar plane. The projected point $\mathbf{X}_1$ is found on line $\mathbf{l}_1$ which is the intersection of the epipolar and projection planes. Lines $\mathbf{l}_1$ and $\mathbf{l}_2$ are called epipolar lines, associated to point **P**. The line from point $\mathbf{O}_1$ to $\mathbf{O}_2$ also intercepts the projection planes. The intersection points $\mathbf{e}_1$ and $\mathbf{e}_2$ are called epipoles. The epipole $\mathbf{e}_2$, is the projection of the optic centre $\mathbf{O}_1$ from the first camera, in the image observed by second one. The epipole $\mathbf{e}_1$ is the projection of $\mathbf{O}_2$ from the second camera onto the image observed by the first.

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are images from the same point, then $\mathbf{X}_2$ must be on the epipolar line associated with $\mathbf{X}_1$. This relation is so called epipolar constraint.

It can be seen, from the epipolar geometry relations, that the single projection point in the first camera corresponds to the epipolar line in the second one. Such coordinates are related by a rigid body transformation, as presented by Yi Ma (2004):

$$\mathbf{X}_2 = \mathbf{R}\mathbf{X}_1 + \mathbf{t} \tag{1}$$

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

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

It can be seen, from equations (1) and (2) 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 (1) can be written in terms of depths and projection coordinates:

$$\lambda_2 \mathbf{x}_2 = \mathbf{R}\lambda_1 \mathbf{x}_1 + \mathbf{t} \tag{3}$$

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

$$\hat{\mathbf{T}} = \begin{bmatrix} 0 & -t_x & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{4}$$

where t_x, t_y and t_z are the coordinates of the translation vector **t**.

Matrix $\hat{\mathbf{T}}$ has the following properties:

$$\hat{\mathbf{T}}\mathbf{t} = \mathbf{0}, \qquad \hat{\mathbf{T}}\mathbf{x} = \mathbf{T} \times \mathbf{x} \tag{5}$$

Premultiplying both sides of equation (3) by $\hat{\mathbf{T}}$ yields:

$$\lambda_2 \hat{\mathbf{T}} x_2 = \hat{\mathbf{T}}\mathbf{R}\lambda_1 \mathbf{x}_1 \tag{6}$$

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

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = 0 \tag{7}$$

where **E** is the essential matrix defined as:

$$\mathbf{E} = \hat{\mathbf{T}}\mathbf{R} \tag{8}$$

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 (7) is also known as the essential constraint relation.

It is assumed that the essential matrix has the form:

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{9}$$

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

$$(E^s)^T = \left[e_{11}\ e_{21}\ e_{31}\ e_{12}\ e_{22}\ e_{32}\ e_{13}\ e_{23}\ e_{33}\right]$$

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

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{19} \\ a_{21} & a_{22} & \cdots & a_{29} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k9} \end{bmatrix} \mathbf{E}^S = 0 \tag{10}$$

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 $\mathbf{x}_2$:

$$\mathbf{a}_k = \left[a_{k1}\ a_{k2} \cdots a_{kj} \cdots a_{k9}\right] = \mathbf{x}_1 \otimes \mathbf{x}_2 \tag{11}$$

A non trivial solution for the homogeneous system of equations is found, for a number k ≥ 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.

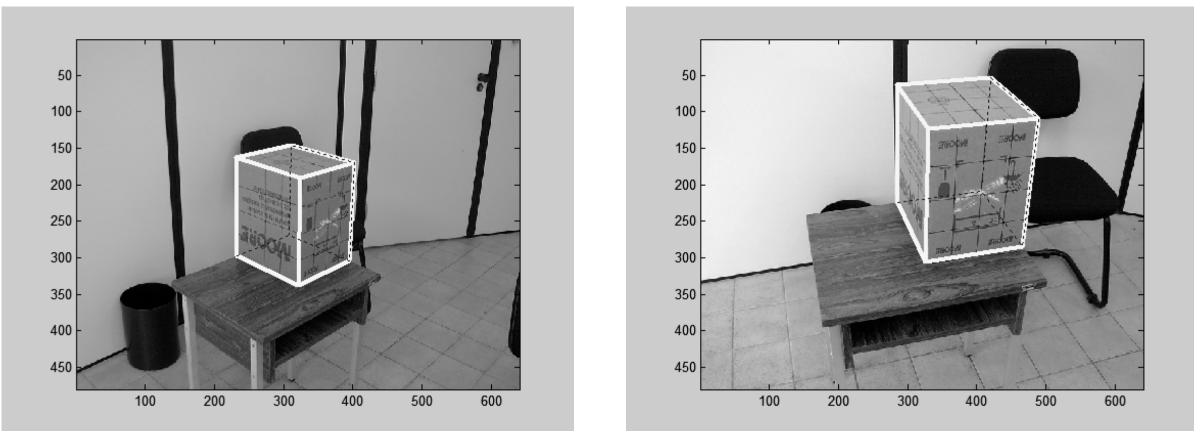
From the estimated essential matrix, it is possible to recover the camera's rotation parameters matrix $\boldsymbol{R}$, and a unit-normalized vector $\boldsymbol{t}$. The module of distance between the cameras origins must be known, in order to perform the estimation of the 3-D position of the projected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

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 $\mathbf{R}$ and the origin translation vector $\mathbf{t}$ are the relative positions of the cameras used for capturing the images.

4. Range Box Method

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 box. A pair of box images are used in the region of interest as presented in figure 2.a and 2.b.



(a) (b)

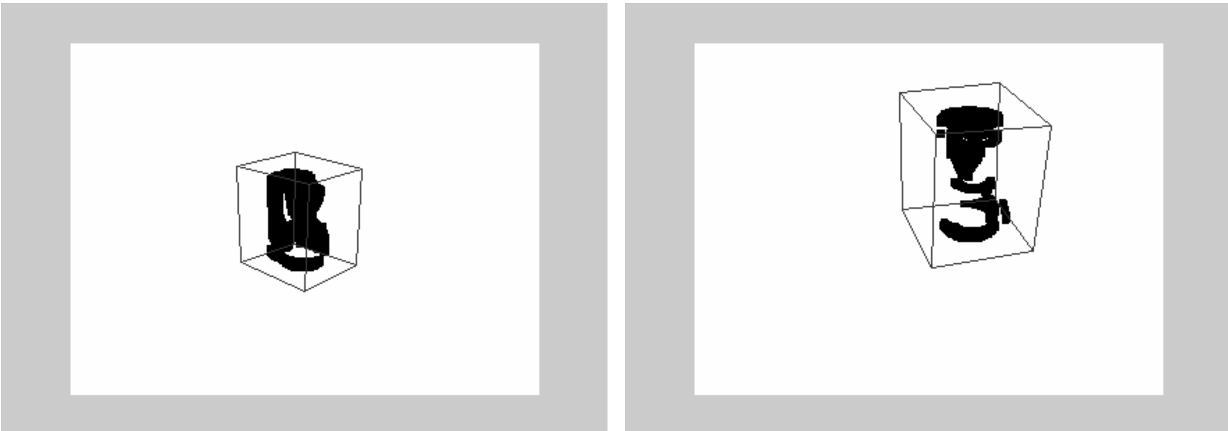
Figure 2. (a) is the left image of solid calibration cube, and (b) is the right image of solid calibration cube.

The figures 2.a and 2.b show two images of a real box used as calibration grid. These images also limit the area around the target object, and are the starting reference for the creation of a virtual box around the object images. Figures

3.a and 3.b show the object in the same space used by the reference box. The figure 4 presents the already segmented objects in the same previously defined starting boxes region.

(a) (b)

Figure 3. (a) Left and (b) right images of the solid.



(a) (b)

Figure 4. Left (a) and right (b) images of the virtual box around a segmented object.

The boxes vertices from images of the figure 2, are obtained by user interaction, as in Y MA (2004). Then the planar image of them are placed on the segmented object image from original images (figure 3.a and b), resulting the new images as showed in the figure 4, a and b.

4.1. Adjusting the Range Box

Adjusting of the range box onto the object's boundary is done in order to map the space occupied by the solid. The process brings each face of the original box to the edges of the segmented object. The algorithm is based on the vanishing points about the perspective of the cube image. Translation of the faces is done in the perspective projections with the aid of vanishing points. External edges of faces of the original box are brought closer to the segmented region, in the perspective views, through rotations about the vanishing points of the box. Each face of the original box is approximated to the segmented region at a time, until all faces are at a minimum distance to the represented object.

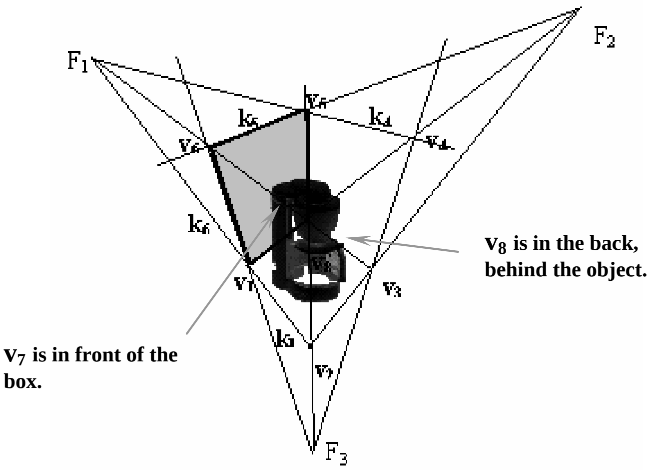
Once the vertices of the new virtual cube are known in both perspective projections, a 3-D reconstruction of the space around the object can be performed. It is important to note that the adjusted box of the first image correspond in an unadjusted box region in the second image, and vice versa. This way, a better estimation of the region that circumscribes the segmented object is that formed by the union of the two spatial boxes that are obtained from adjustments of the original reference figure in both perspective views.

A summary of steps for application of the proposed technique is listed below:

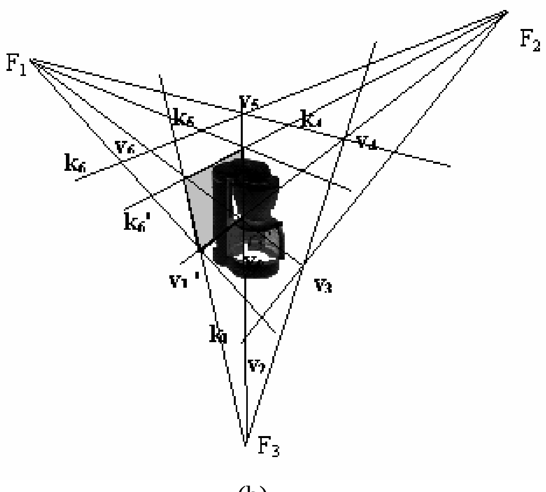
The proposed method uses the reconstruction algorithm from Yi Ma (2004) and it is how follow:

I. Find a correspondence of points of the projected images for estimating the essential matrix, or camera's pose. A calibrating box can be used for that purpose, with the advantage that its vertices can also serve as points of the starting virtual range box. The eight point algorithm can be used to perform such estimation.
II. Perform a segmentation of the region of interest, where lies the solid whose dimensions are to be approximated.
III. Approximate the boundaries of the starting range boxes to the region of interest through rotations about the vanishing points of perspective representation of the virtual box.
IV. Compute the spatial coordinates of the vertices of the adjusted virtual range boxes, defining the approximate dimensions and position of the solid.

The figure 5.a presents the initial process of adjusting the range box to the object. The range box is formed by vertices v_n, with n= 1,2,...8. Vertices v_7 and v_8 are unknown and must be found by intersection of the known lines k_n, where n= 1,2,...6. The vanishing points F_1, F_2 e F_3, can be estimated in each view, from the intersection of parallel edges of the projected box. The vertex v_7 is obtained by the mean intersection points: $F_1v_6 \cap F_2v_4$, $F_1v_6 \cap F_3v_2$ and $F_2v_4 \cap F_3v_2$. In a similar way, point v_8, is found from the mean intersection points: $F_1v_3 \cap F_2v_1$, $F_1v_3 \cap F_3v_5$ and $F_2v_1 \cap F_3v_5$. The bounding box then formed by eight known vertices, that is $[v_1\ v_2\ v_3\ v_4\ v_5\ v_6\ v_7\ v_8]^T$.



(a)

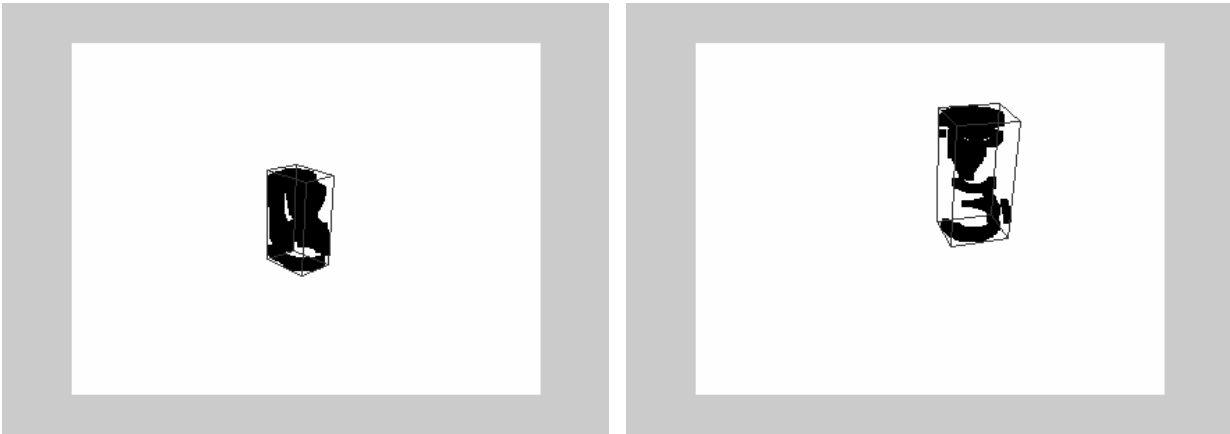


(b)

Figure 5. (a) The initial bounding box around the object.
(b) The adjustment for the face formed by vertices v_1, v_8, v_5 and v_6.

The face with vertices v_1, v_8, v_5 and v_6 of the bounding box is approximated to the edge of the object, as presented in the figure 5.b. This approximation is achieved by the rotation of the border lines k_5 e k_6, with pivots in F_3 and F_1

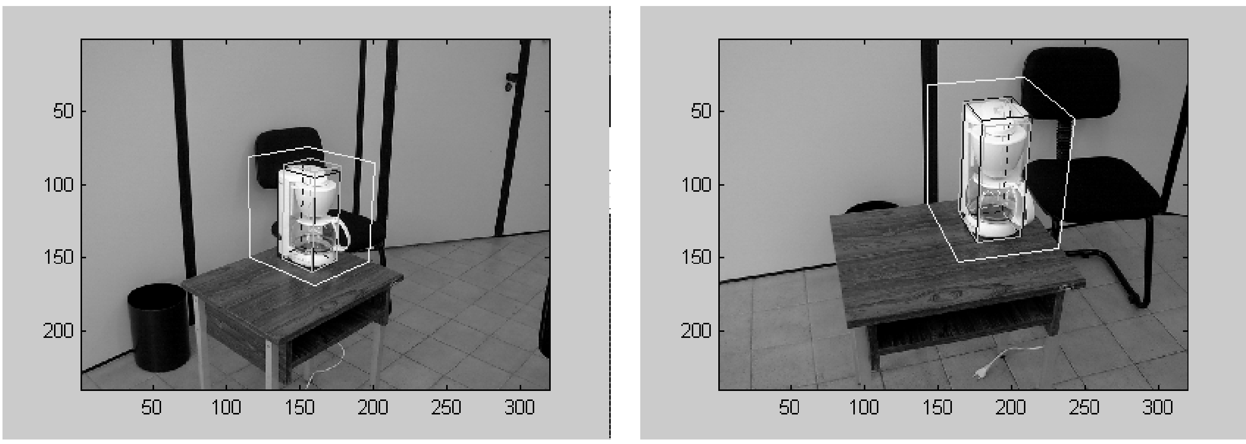
respectively, until the first of them touches a point (pixel) in the region of interest. In this situation, new vertices v'_1, v'_6, and v'_5 are determined by the intersection of two rotated edges with the original, not rotated edge. The approximation scheme is repeated to all of the remaining box faces.



(a) (b)

Figure 6. Left (a) and right (b) images of adjusted box projections.

The process is executed for the right and left images. The figure 6 shows the two resulting images. They are used in a 3-D reconstruction process. A simplified spatial model from the union of the boundary boxes is built, where position and dimensioning information is represented by the calculated center of gravity and principal axis of inertia of the resulting box. The final 3-D range box, is achieved by uniting the two boxes obtained on each projection. Figure 7 presents the range boxes adjusted to the object.



(a) (b)

Figure 7. External starting virtual box and adjusted boxes form left and right image projections.

The work by Chan (2005), shows that in 3-D visualization schemes, the reconstruction of surfaces require large quantities of points (usually referred as “cloud of points”). The results in such an approach is a model with a large number of facets and vertices, which demands substantial computational effort. In this context, a method is proposed to map the space around an object, using a virtual box. The virtual box can be used to approach the object borders. Volume and position can be then estimated in an approximated form. The virtual box of reference, also called “bounding box” is present at some different applications. It can be seen in Chu (2001), which presents a method of 3-D model reconstruction based in a reference cube and strips of light. (LDP - Light Plane Definition Points). This concept can be also used for image indexing as in Barequet (2001), volume and positioning as in Kurka (2005), packaging in Chan (2001), design for Assembly – DFA in Coma (2003), CAD models and rendering in Chan (2005).

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 method proposed here is to be applied in similar situations where a pair of stationary 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 same technique of circumscribing boxes can be used in applications where measurement precision is a requirement, by subdividing the initial box of iteration in a number of smaller starting boxes.

5. Conclusions

The paper displays the progress of studies to estimate the position and dimensions of a solid in a manufacturing cell. The background to image processing theory is presented, including the form of the essential matrix, which is the base for 3-D reconstruction techniques. Practical use of the 3-D reconstruction from a pair of images is proposed by means of definition of a virtual box that is fitted 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 region of measuring interest. The method represents a rather fast way of defining the 3-D region occupied by a solid, without the need to perform a complete surface/volume reconstruction. The method can be applied to image based positioning for a mobile robot, or tracking movement of persons or objects. The work is under progress and the first practical results are being obtained.

6. References

Barequet, G.; Har-Peled, S. , 2001. “Efficiently Appoximating the Minimum-Volume Bounding Box of a Point Set in Three Dimensions”; Journal of Algorithms, n.38; pp. 91-109.

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.

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.

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

Chan, C.K.; Tan, S.T., 2001. “Determination of the minimum bounding box of an arbitrary solid: an iteractive approach”; Computer and Structures, n.79, pp.1433-1449.

Chan, C.K.; Tan, S.T., 2005. “Extreme points of a large 3D point set along multiples directions”; Computer-Aided Design, n.37, pp.17-34.

Chu, C. W.; Hwang, S.; Jung, S. K., 2001. "Calibration-free Approach to 3D Reconstruction Using Light Stripe Projections on a Cube Frame."; IEEE, pp.13-19.

Coma, O.; Mascle, C., Véron, P., 2003. “Geometric and form feature recognition tools applied to a design for assembly methodology”; Computer-Aided Design, n.35, PP.1193-1210.

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.

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

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

Kurka, P. G.; Rudek, M; Menezes Jr, L.F., 2005. "3-D Image Signal Processing for Automated Operations Using a Range Cube". Proceedings of the XI DINAME, Ouro Preto - MG - Brazil

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.

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

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.

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

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

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

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

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

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.

7. Responsibility notice