ABCM Symposium Series in Mechatronics - Vol. 4 - pp.38-43


# MULTIPLE 3D OBJECTS IDENTIFICATION FROM IMAGES BASED ON BOUNDING BOX RECONSTRUCTION

**Marcelo Rudek, marcelo.rudek@pucpr.br**
**Osiris Canciglieri Jr, osiris.canciglieri@pucpr.br**
Pontifícia Universidade Católica do Paraná – Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PUCPR/PPGEPS) – Rua Imaculada Conceição, 1155 - Prado Velho - Curitiba/PR
CEP: 80215-901

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

***Abstract.*** *The work shows a procedure to determine the information of 3D position of multiple solid objects in the working space of a robotic cell. The proposed method uses a pair of images to obtain a 3-D mapping of the space around the objects. A calibration box is initially positioned to define a reference frame, and from it is constructed a virtual axis aligned bounding box. This 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 these adjusted boxes. The method represents a rather fast way of defining the 3D region occupied by a solid, without the need to perform a complete surface reconstruction. The proposed technique has applications in robotics, automation and manufacturing.*

***Keywords****: computer vision, 3D reconstruction, image processing, bounding box, robotics*

## 1. INTRODUCTION

Three-dimensional reconstructions are used for various purposes in cases where the depth of information is essential to understanding a scene. A scene in a digital image is composed of a complex structure of various elements, which are not always required in a reconstruction. Some key points of the environment may be sufficient to determine the movement of a robot, identify the presence of an object and retrieve its basic geometrical characteristics. The simplifying of the environment can reduce the computational effort by the suppression of information that is not necessary to analyze a scene. In many cases, the decrease in the amount of information can improve the efficiency of the process without compromising the end result.

The knowledge about object's positions is a very important question to define the robot position in automated manufacturing cells. The construction of a virtual bounding box can simplify the amount of parameters used in the computational 3D coordinate's calculation. Many theories, techniques, mathematical and computational tools are being developed for recovering of 3D information from image analysis. The use of bounding box concept to determine the position and volumetric information of a 3D object is an important tool to real world mapping. The idea of the use of bounding boxes was initially presented by Gottschalk (2000) to identify a collision of objects that are moving. Some other examples of application of this can be seen in Barequet (2001) to indexing and retrieval of images, in Chan (2001) for processes of packaging, in Shin (2004) for representation of surfaces, in Coma (2003) to design for assembly (DFA), in Majchrzak (2004) to perform analysis of the movement of objects in two-dimensional scenes, and in Chan (2005) to simplify the representation of a 3D object from a CAD model, among others. Some studies about the bounding box concept have been development by kurka (2005) and Rudek (2006). The approach consists in the fact of the information of range can be achieved by inscribing the object in a virtual bounding box. It is created from known vertices of the one reference real box. The vanishing points obtained of a pair of stereo images of the reference box it's the starting procedure to the face box adjustment. The box is adjusted around the object's image projection border through perspective translation of its faces to fit a region of interest. The volumes and positions defined by the adjusted boxes are approximations for the real dimensions and location of the solid. 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 way of defining the 3D region occupied by a solid, without the need to perform a complete surface or volume reconstruction as presented in Kurka (2007).

This paper shows the possibility of the use of this method now for multiple object identification. The proposal is to adapt the method of bounding box used in the one object identification, to the new approach to use within environment that contains multiple objects. The present work also considers the pinhole camera configuration, in order to determine geometric aspects of objects and its position and dimension. It presents an overview of mathematical model of a camera, as well as calibration process and a brief description of the epipolar geometry relations that are the basis of 3D image reconstruction. It is also presents a case of 3D position identification of the cylindrical objects in a real integrated manufacturing cell operating by the robot guided by a vision system.

ABCM Symposium Series in Mechatronics - Vol. 4 - pp.38-43


## 2. BACKGROUND CONCEPTS

The computing of position of one object in space requires mathematic equations that link the 3D regions with their corresponding image point projection from calibrated camera parameters. The 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 3D information from a pair of 2D 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. The cameras positions are represented by their rotation (**R**) and translation (**t**) matrices. The mathematical representation of an image is achieved through a relationship model between the world coordinates of a 3D point and its corresponding position in the image plane. Such a relationship, which comes from the pinhole model, is described in Forsyth (2003). Establishing correspondences of points between two images however is a difficult task. Then, the epipolar constraint geometry described in Ma (2004), helps to limit the searching regions for matching points. The relationship between two image views $\mathbf{X}_1$ and $\mathbf{X}_2$ of the same point is given by the rigid body transformation, as presented in the Eq. (1), as:

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

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 where relative camera rotation **R** and origin translation vector **t** are not a priori known. The estimation of such quantities is the starting point of 3-D reconstruction techniques.

### 2.1. Epipolar Geometry

When the intrinsic parameters of camera are known, obtained by a calibration process, the epipolar transformation is represented by the essential matrix, **E**, comprised of the parameters of rotation and translation of the camera reference systems, as Eq. (2) shows. Otherwise, for uncalibrated system, the epipolar constraint is given by the fundamental matrix, called **F** as in Eq. (3) where **K** matrix is called the intrinsic parameters, or calibration matrix.

$$\mathbf{E} = \hat{\mathbf{T}}\mathbf{R} \tag{2}$$

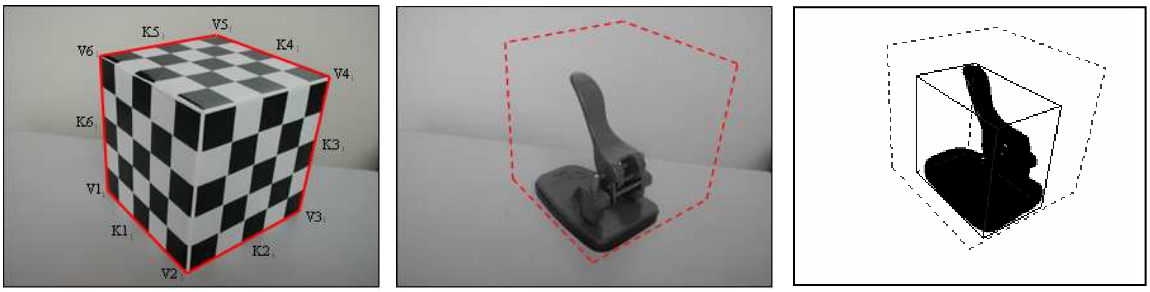
$$\mathbf{F} = \mathbf{K}^{-\mathbf{T}}\hat{\mathbf{T}}\mathbf{R}\mathbf{K}^{-\mathbf{1}} \tag{3}$$

The formation of these matrices is widely discussed in Ma (2004). For this study the intrinsic parameters are estimated by calibration, and the relative pose between two cameras is calculated using the essential matrix.

### 2.2. Axis Aligned Bounding Box (AABB)

Two types of bounding boxes are described by Gottschalk (2000). One is called AABB (Axis Aligned Bounding Box) that is a parallelogram whose faces can be represented as the extension of lines over a frame of reference. In this case the adjustment requires that the object is geometrically positioned according to a known reference box of initial approximation (outer box). Another way is to calculate the fit through the type called OBB (Oriented Bounding Box) that can be constructed using principal component analysis (PCA) or calculating the inertial moments of the objects.

In the method proposed by Kurka (2005) a real reference box is used to initial configuration as shown in Fig. 1.a. This same box is also used in the calibration process. Then it is realized an adjustment processes based on vanishing points as described in previous paper of Kurka (2007), and it is obtained a virtual reference box as in Fig. 1.b. The result is presented in Fig. 1.c, where doted line is the initial box and the solid line is the approximation, which represents the volume. The example in Fig.1 it is a brief of the method, and the all detailed process is described in Rudek (2006).




(a) (b) (c)

Figure 1. Example of an AABB faces adjust for only one view of an object.

ABCM Symposium Series in Mechatronics - Vol. 4 - pp.38-43


## 3. PROPOSED METHOD

The method proposed is an extension of traditional method described in Kurka (2005). The same manner an AABB can be adjusted on the solid's edges for multiple objects, as shown in the Fig. 2.

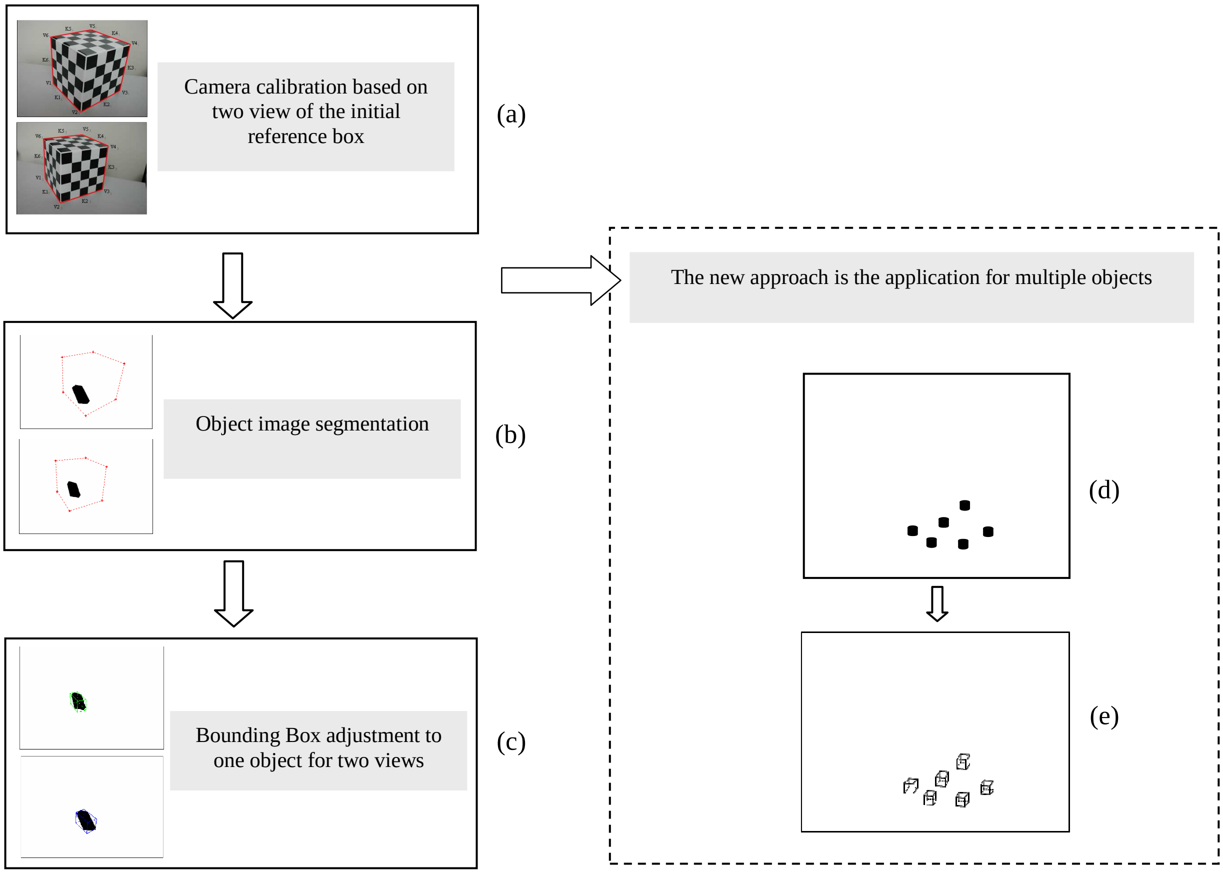



Figure 2. Sketch of a method for multiple objects positioning estimation.

As presented in Fig. 2.a, the calibration pattern is the same for the use to one or more objects. It's assumed that the camera do not changes the position. This box forms the reference initial AABB. The following step in Fig. 2.b, deals about the inserting the object in the reference region, and the capture of a stereo pair of images of the object. Then the tresholding must be resolved to obtain the ROI (Region of the Interest). To simplify the object in terms of pixels quantities, the edge identification is a good alternative. For this kind of application, the gradient based algorithms are sufficient to edge locations even if it generates nonclosed contours. The gradient definition is based on the first-order derivative of image, where it is zero in the regions with constant values of black or white, but is not zero when calculated on a discontinuity. The first-order derivatives are based on 2D gradient. Some gradients operators as Sobel or Canny (Gonzalez, 2002) are frequently used to edge detection.

For the next step represented by the Fig. 2.c, the object must exist into box space, as indicated with the line around the object space. Then, in this step performs the AABB adjustment trough the faces movement on the object borders.

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

i. Use Find a correspondence of points of the projected images for estimating the essential matrix. 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 bounding boxes, defining the approximate dimensions and position of the solid.

ABCM Symposium Series in Mechatronics - Vol. 4 - pp.38-43


The idea is to use this method for more than one object. Then for this new proposal, the steps ii, iii and iv, can be repeated for each object with some modifications. The initial problem is also to identify de ROI for each one. Each solid, after image segmentation and identification, must be adjusted separated, because the vanishing point used in the process must be evaluated separately too. The center's positions of objects can be found using a morphologic black and white operation called "erode", then the centers positions can be located and quantities can be calculated in 2D segmented image. The objects are again reconstructed from local marks trough a "dilatation" process and the result is as in Fig. 2.d. The adjusted bounding boxes represented in Fig. 2.e, can obtained by the translation of reference frame of outer box, to known coordinates of each solid segmented image. Then the adjustment can be performed for each solid, one by one. The new steps (Fig. 2.d e Fig. 2.e) are the objective of this proposed method. The Fig. 3 shows the vertices identification for a demonstration model.

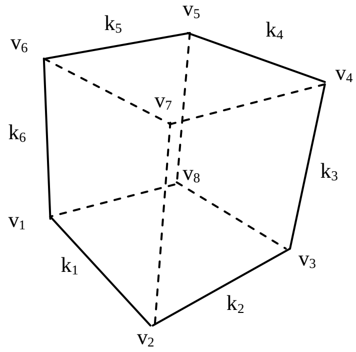
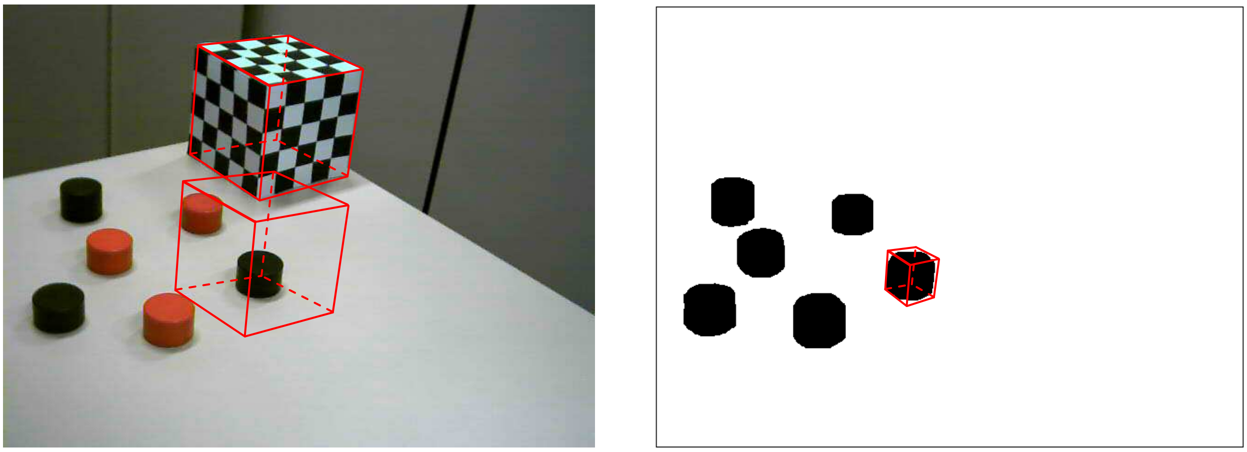



Figure 3. Virtual box vertices identification.

From the calibration rig captured by the camera, it can build a virtual box, based on vertices positions. The extraction of the vertices named $v_i$, $i=1\ldots6$, of the box contour, can be by user interaction. The $v_7$ and $v_8$ can be calculated based on the vanishing points estimation. The camera must see at least six border lines (solid lines identified by $k_n$ in Fig. 3) that are needed to obtain the vanishing point as described in kurka (2005) and Rudek (2006). The virtual cube formed in the space can be reprojected on the 2D space by perspective projection on the image plane. In this case is not necessary to have an object occupying the area into the reference box location. The Fig. 4.a presents an example, where de calibration cube was used to the vertices and cube lines identifications, and its projection is changed for a specific piece. It is represented only one image of the scene, but the process is based a stereo pair of images. Then instead perform the adjustment to an object within box region, these vertices values are moved to known pieces positions, and the adjustment is made for each, as in Fig. 4.b.



(a) (b)

Figure 4. Example of bounding box identification.
(a) original image with calibration cube. (b) segmented objects and adjust for one piece.

The hidden box vertex $v_8$ (presented in Fig. 3) is moved to the central coordinate of target. This guarantees that six external cube lines are visible. It is observed that exist a distortion caused by the translation, because the perspective changes. This problem will be described in next section. To evaluate the performance of the method it is presented a case study based on real application in manufacturing.

ABCM Symposium Series in Mechatronics - Vol. 4 - pp.38-43


## 4. A CASE STUDY

A case of study was implemented from an initial work developed by Rocco (2007) where a computer vision system operates in a robotic manufacturing cell. This system can find the cylindrical objects positioning based on a top 2D view of a bench by the camera mounted on the manipulator. He proposes an algorithm that verifies whether the robot can access an object, and perform the best way to catch the solid as presented in Fig. 5. Despite the system works to bi-dimensional analysis, the side view of solid can't be determinate. So, the height isn't evaluated.

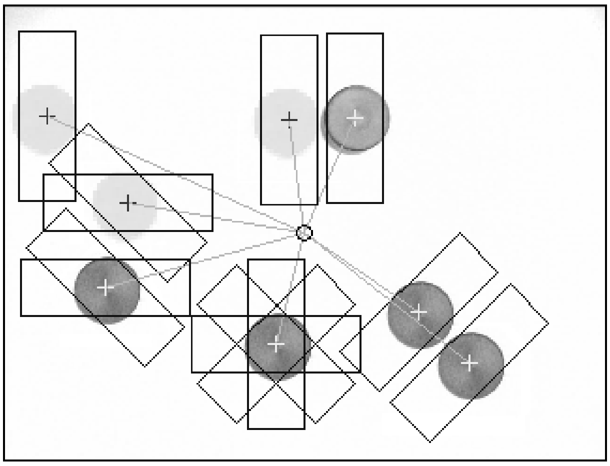


Figure 5. Example of the 2D objects position identification and the calculated positions to manipulation.

Then the idea is to perform a new function in this system that deals of volume's analysis. In this case, the image capture system was positioned out of robot manipulator. The camera position is in any position able to view all the workspace. The calibration process used was how defined by Zhang (2000) and it is based on matching algorithm created by Harris and Stephens (Harris, 1998). The initial condition it's to estimate de essential matrix according as presented in the item 2.1 using a calibration pattern (calibration cube). How the points of reference cube is now known, this box is created on 3D space based on the reconstruction.

The cube vertices are known and reference frame now can be moved for each object image trough the known centers. It is considerate that the camera positions do not change and the objects positions do not change. The robot will find each object in the point $P_k=[x_k,y_k,z_k]$ coordinates searched by segmentation process, where $x_k$ and $y_k$ are coordinate in the referenced plane XY of object's centers of solid $k=1,2,\ldots,N$ and the height of each one is given by Z orthogonal direction and will define frame movement to catch the object. The Fig. 6 presents a result of this process.

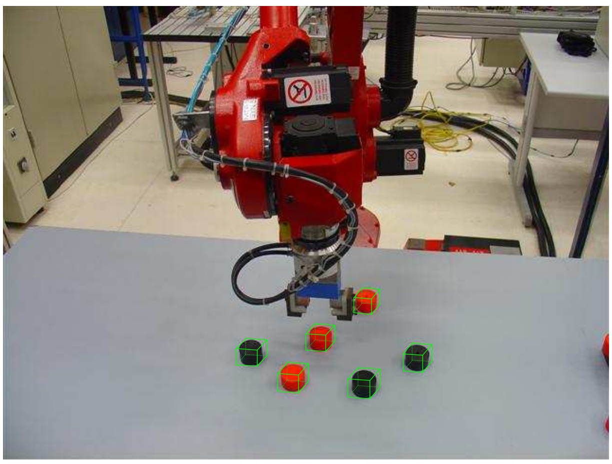


Figure 6. Example of bounding box identification in a manufacture cell.

To obtain the multiple bounding boxes as shown in Fig. 6, was developed a C++ program to capture the images. The calibration process was developed using MatLab because the matrix manipulation is totally implemented in the specific toolboxes. The calibration parameters are passed to C++ program by an ASCII file, which work about the projective scene information, and basically perform a segmentation algorithm. After the program look for the centers and draw the adjusted box lines. The boxes lines presented on the objects are the projection of the reference virtual cube. In this point it is important to remember that the vertices of calibration real cube are projected into 3D space, and after this, these same vertices are reprojected onto 2D image plane.

The original algorithm works considering that the object always into the same space occupied by reference cube, as in Fig.1. In this study, the one or more objects are out the space occupied by reference box. When the AABB is formed, it necessary changes its position. How the objects are closer of cameras than original cube, the perspective is distorted, as can be seen in the Fig. 4.a. In the case study realized, was observed that the height of each AABB generated is close enough with the real object height. Also, there is another problem concerned with the upper visible face of object that is confused as height according the perspective. Then the boxes generated are slightly higher. This problem was expected, and the error can be compensated in the robot movements program.

## 5. CONCLUSION

Despite it is a rough procedure for some cases where the object is sparse in the space, the method offers guaranties about the position and their volume representation. The 2D centre coordinates can permit the correct manipulator positioning, and the height is guaranteed by the volume of adjusted box. In the example of this work, for each specific object position, the different sizes of the boxes can be recovered, and then the volumes will be different. Then the top down robot's movement can be adjusted based on this volume value. The axis aligned bounding box can be used, since the task does not require a great precision, because the calculated volume defines only the initial approximation for the movement. Once calibrated, while the camera pose doesn't change, the process can be executed for all objects without new calibrations. Whether a new type of object is placed, and it smaller than reference box, the system also works. This study is in progress yet, and the future objective is improves the segmentation method to use only the side view of object, to perform a better adjustment.

## 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.

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.

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.

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

Gonzalez, Rafael C. 2002; Woods, Richard E. "Digital Image Processing, 2/E". Prentice Hall.

Gottschalk, S.; 2000. "Colision Queries Using Oriented Bounding Box". Phd Thesis, Univ. of North Carolina, USA.

Harris, C. and Stephens, M., 1998. "A Combined Corner and Edge Detector". Proceedings of the 4 Alvey Vision Conference, Manchester, England, pp. 147-151.

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

Kurka, P. R. G.; Rudek, M.; 2007. "3-D Volume and Position Recovering Using a Virtual Reference Box". IEEE Transactions on Image Processing, pp.573-576.

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.

Majchrzak, D.; Sarkar, S., 2004. "Parallelizing Motion Segmentation by Perceptual Organization of XYT". IEEE-Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04).

Rocco, T. F.; 2007. "Desenvolvimento de um Sistema Baseado em Imagens para Identificação de Peças Manipuladas por Braço Robótico". Pontifícia Universidade Católica do Paraná, Brasil.

Rudek, M.;2006."3D Positioning and Dimensioning Method Based on Digital Images".Doctoral thesis, Unicamp, Brasil.

Shin, D.; Kurfess, T. R., 2004. "Three-dimensional Metrology of Surface Extracted From a Cloud of Measured points". Precision Engineering, n.28, pp.302-313.

Zhang, Z., 2000. "A Flexible New Technique for Camera Calibration". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22. N.11, pp.1330-1334.


## 7. RESPONSIBILITY NOTICE

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