AN EYE-IN-HAND ROBOTIC VISION SYSTEM FOR RECOGNITION AND LOCATION OF OBJECTS BASED ON GEOMETRIC INVARIANTS

José Maurício S. T. Motta, jmmotta@unb.br
Laércio Arraes Jardim, laerciojardim@brturbo.com.br
University of Brasilia – Faculty of Technology, Department of Mechanical Engineering, Brasilia-DF, Brazil, 70910-900

Abstract. *Robotic vision automatic systems for recognition and location of industrial objects still present several difficulties, due to variations in illumination and complexity of geometrical shapes in industrial environment, including specular reflection in surfaces and algorithms of high computational cost. This article presents a technique for recognition and location of objects in industrial environments using an algorithm based on the well-known geometric invariants, with extraction of corners (feature points) and use of specific invariants for points and lines under projective transformations. The computational cost is substantially reduced by using an ordered matching routine around the perimeter of the binarized object image. The bank of invariants is comprised by bi-dimensional images previously acquired of the object-models, and experimental tests had demonstrated high degree of effectiveness in the recognition of these objects. Object location relative to the camera coordinate system allows robotic manipulators to grasp objects with position control and orientation. Approaches using neural networks for determination of corners and for binarization of the images are also being developed to reduce the restrictions of positioning and illumination.*

***Keywords**: Robotic Vision Systems, Object Recognition, Geometric Invariants, Object Location*

1. INTRODUCTION

One of the main research areas in robotic vision today is to make feasible to a robot to recognize and understand the objects in sight. Some researchers have proposed stereo vision techniques and high level knowledge structured processing models (Kuvich, 2005, Sumi *et al*., 2004), however the development and implementation of such techniques have been slow, mainly due the need to process high complex tasks required in image analysis, high cost and low computational speed, reducing the precision and the flexibility of such systems.

In general, object recognition is performed by extracting important features from the image, such as points, straight lines, curves, following by matching these features to those of the object models previously stored. However, perspective transforms, which occur in 2-D imaging, produce different images according to the camera viewpoint, turning the process of comparing image to model features a difficult task. Some complicate high-level techniques propose to reduce this undesirable effect in projective transforms (Lu *et al*., 2003, Basak, 2000). Another difficulty in object recognition is to make available object models that can be conveniently compared to the object images, which means that important features of the scenes have to be found in the models, no matter how the object is being imaged. The object models database have to be the most complete as possible.

However, geometric invariants (Zongmin *et al*., 2005, Löpez-Franco, 2006) are known to provide much easier solutions to the 2-D perspective imaging problems, which have properties that are invariant to the viewpoints. This advantage is quite clear in the case of industrial objects in scene, since they have usually well defined straight contours and planar shapes.

This article presents a technique for recognition and location of objects in industrial environments using an algorithm for extraction of corners (feature points) and specific invariants for points and lines under projective transformations. A technique to locate objects without the need to calibrate the camera parameters is also presented.

2. CORNER EXTRACTION

There are several techniques available to extract corners from an image (Trucco and Verri, 1998 e Noble, 1988). In this work it will be used the image spatial gradient $[E_x, E_y]^T$ to find such corners, calculated in two orthogonal directions. The corner is extracted from the smallest eigenvalue of a strength matrix C, calculated in a neighborhood Q (2N+1 x 2N+1), in image coordinates (pixels), of a point *p*, defined as (Trucco and Verri, 1998):

$$C = \begin{bmatrix} \sum E_x^2 & \sum E_x E_y \\ \sum E_x E_y & \sum E_y^2 \end{bmatrix} \quad \text{where} \quad E_x = \frac{\partial E}{\partial x} \quad \text{e} \quad E_y = \frac{\partial E}{\partial y} \tag{1}$$

A threshold value, τ, can be used as a fraction of the largest values amongst the smallest eigenvalues of C. This assures that in two directions from a point the gradient values are large enough to define a corner. The result is an increasing list of corners ordered by their location along the horizontal lines of the image.

3. INVARIANT PROJECTIVES TO CONSTRUCT MODELS

Invariants are functional properties of geometric configurations that do not change under certain classes of transforms, for example, the length of a line segment does not vary as a function of a rotation or translation. In the usual transforms, there are well-known invariants as lengths and areas as in Euclidean transforms (Vicente *et al.*, 2002), or ratio of lengths and angles as in transforms by similarities (Hartley and Zisserman, 2000). In projective transforms there are also invariants, and the most known is the cross-ratio.

In this section, projective invariants constructed with image points are described. These invariants are calculated using 5 corner points of the image that are not collinear (p_i $i= 1,2,3,4,5$) in homogeneous coordinates. A point in a 3-D object can be represented in homogeneous coordinates as P = $(X_i,Y_i,Z_i.1)$ and in an image as $p_i = (x_i,y_i,1)$. Several authors (Trucco and Verri, 1998, Bong, 2000 and Vicente *et al.*, 2002) have used two independent projective invariants I_1 and I_2 defined as:

$$I_1 = \frac{S_{431} \cdot S_{521}}{S_{421} \cdot S_{531}} \qquad I_2 = \frac{S_{421} \cdot S_{532}}{S_{432} \cdot S_{521}} \tag{2}$$

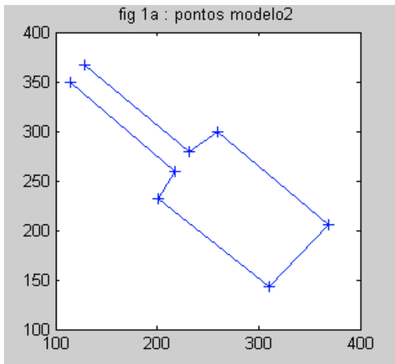
, where S_{ijk} is the determinant of the 3x3 matrix defined by the points $|\, p_i \quad p_j \quad p_k \,|$.

3.1. Object Modeling

The object models can be constructed using the algorithm to extract corners, described in section 2, on the object image, and a set of these corners is selected such as the determinant of S_{ijk} is not null, which prevents undesirable values for I_1 and I_2 (null or undefined). This requires that three points amongst five are not collinear.

In this experiment two object models had been used, showed in Fig. (1) and Fig. (2). These objects were chosen because of the well defined contour of the part, similar to common industrial objects.

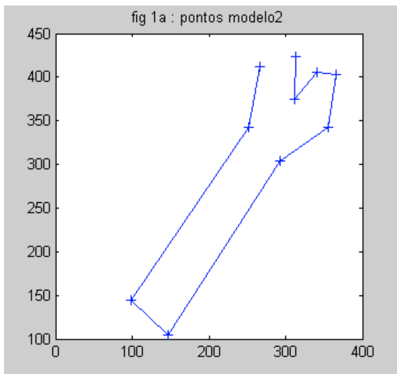
(a)



(b)

Figure 1. Model 1 image. (a) Model 1 image with corners (b) Model 1 corners

(a)



(b)

Figure 2. Model 2 image. (a) Model 2 image with corners (b) Model 2 corners

The sequence of points selected from the set of corners extracted from the image must follow an ordered sequence such that they form a closed polygon in the clockwise direction, as it is shown in Fig. (3). In Fig. (1-b) and (2-b) the corners are also ordered in the clockwise direction.

The objective of ordering the set of selected corner points in a sequence in the clockwise direction is to reduce the number of possible combinations to identify the object. It is suggested that the image is binarized before applying the algorithm to select the corner points. The algorithm is initiated with a point *p* belonging to the image contour and subsequently the search window is moved in the clockwise direction along the entire object contour, generating a list of corner points as long as they are found.

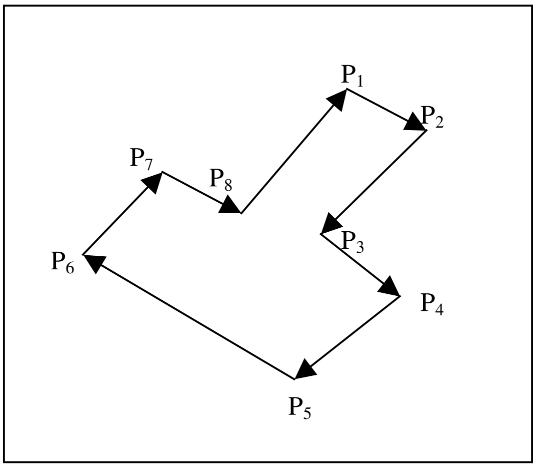


Figure 3. Ordered sequence of corners in the clockwise direction.

Assuming that the model image contains M corners, two vectors of invariants *g1* and *g2* are defined to represent the model as (Bong, 2000):

$$g_i = \begin{bmatrix} I_i(1,2,3,4,5) \\ I_i(2,3,4,5,6) \\ I_i(3,4,5,6,7) \\ \\ \\ I_i(M-1,M,1,2,3) \\ I_i(M,1,2,3,4) \end{bmatrix} \qquad i = 1,2 \tag{3}$$

, where I_i are invariants of the corners calculated with Eq. (2).

4. OBJECT RECOGNITION

Object recognition requires that the geometry of these objects is known previously. There are no means how to recognize an unknown object, that is to say, object features have to be previously recorded.

The central idea of the method is in the comparison between a 2-D image and a database of known object models. The technique for object recognition by means of geometric invariants developed in this work requires two types of routines that are shown in the flowchart of the Fig. (4).

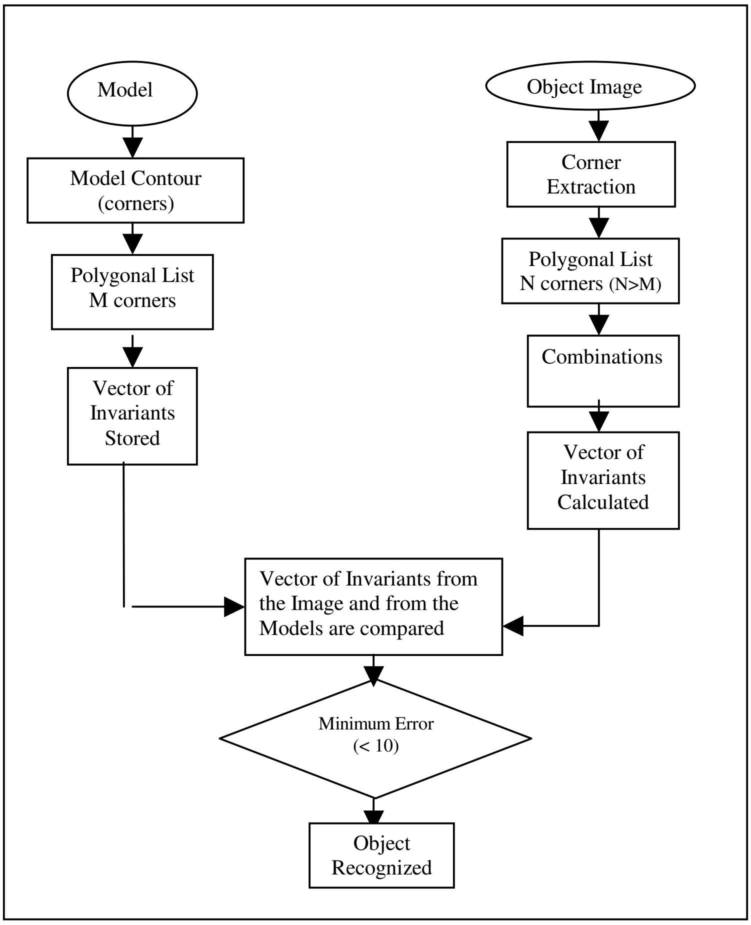


Figure 4. Routines for object recognition by using geometric invariants

The first routine (off-line) has the objective of constructing a bank of invariants, where each invariant vector represents an object model. For the construction of this bank it is necessary to identify the geometry of the part (contours), to make a list of feature points (M corners) and finally to generate the vector of invariants using Eq. (3). The second routine (on-line) has the objective of identifying an object in an image. The routine is initiated with the extraction of corners (item 2) to make a list of a polygonal set of points (N corners) that geometrically represents the object to be identified. Subsequently, the vector of invariants is calculated for each combination (${}_{N}C_{M}$; N>M) and compared with all the models stored by the off-line routine. The object can be identified from the minimum error, defined as the smallest difference between the calculated invariants in the image and in the model. The error can be calculated from Eq. (4):

$$E = \sum_{k=1}^{M}\left(\left|\frac{g_{1k} - \hat{g}_{1k}}{g_{1k}}\right| + \left|\frac{g_{2k} - \hat{g}_{2k}}{g_{2k}}\right|\right) \tag{4}$$

, where g_{ik} is the kth entry of the g_i vector and $\hat{g}_{ik}$ is the invariant correspondent to g_{ik} in the image.

5. OBJECT LOCATION

Generally, in stereoscopy two or more well calibrated cameras (intrinsic and extrinsic parameters have to be known) are necessary to locate an object in images. When these camera parameters are unknown, the problem of object location becomes more complex. When the camera parameters are known, even with a single image (camera) it is still possible to locate parts of 3-D objects using specific techniques for location. However, the problem with monocular vision becomes highly more complex if these parameters are unknown. As an alternative to by-pass the difficulties involved in object location with monocular vision it is proposed here a technique that uses the concept of perspective transforms.

The method is based on perspective transforms that preserve cross-ratios of 4 collinear points and the relation between 2-D coordinates (x,y) of a point in the object plane and its perspective projection (x', y') in the image plane. This relation can be written as (Horaud *et al.*, 1989):

$$x' = \lambda_1 \frac{ax + by + 1}{px + qy + 1}$$
$$y' = \lambda_1 \frac{cx + dy + 1}{px + qy + 1} \quad (5)$$

, where a, b, c, d, p, q, λ_1 and λ_2 are constants.

The equations

$$x' = \frac{u}{w} \;,\; y' = \frac{v}{w} \;,\; a' = \lambda_1 . a \;,\; b' = \lambda_1 . b \;,\; c' = \lambda_2 . c \;,\; d' = \lambda_2 . d \quad (6)$$

together with Eq. (5), can be rewritten to produce Eq. (7), which is very similar to the basic equation of the perspective projection in the image plane:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} a' & b' & \lambda_1 \\ c' & d' & \lambda_2 \\ p & q & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

In fact, one has to determine (x', y') in the object plane from the camera coordinates (pixels). For that, the coefficients a', b ', c ', d ', p, q, λ_1 and λ_2 of the matrix in Eq. (7) must be determined.

The equation above represents a system of 2 equations with 8 coefficients that must be identified, needing at least 4 not collinear points in the image plan (x ', y ') and their correspondents in the object plan (x/z, y/z, 1) to solve the system. This linear system results in the equation $A.x = b$ and can be written as in Eq. (8).

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x_1' & -y_1 x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y_1' & -y_1 y_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2 x_2' & -y_2 x_2' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2 y_2' & -y_2 y_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3 x_3' & -y_3 x_3' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3 y_3' & -y_3 y_3' \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4 x_4' & -y_4 x_4' \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4 y_4' & -y_4 y_4' \end{bmatrix} \begin{bmatrix} a' \\ b' \\ \lambda_1 \\ c' \\ d' \\ \lambda_2 \\ p \\ q \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4' \end{bmatrix} \quad (8)$$

Using the least square method , the solution of the above system can be obtained through the following equation:

$$A^t A x = A^t b \quad (9)$$

Thus, any point on an object in 3-D space can be located from the coordinates of this object in the image.

6. EXPERIMENTS AND RESULTS

Three types of experiments had been carried out:

a) experiments to extract sets of corners from the object images and store invariants of the model images;
b) experiments to compare images of the object models to the stored bank of invariants using the minimum error in Eq. (4);
c) experiments for object location in an image, with the purpose of handling these objects.

The results are presented in the Fig. (5), (6), (7) and (8) and in Tables (1) e (2):

Figure (5) shows an object and various corners. In Table (1) the corners are ordered in an increasing sequence in the horizontal lines of the image. In Table. (2) these same corners are ordered forming a closed polygon.

Table 1: Corner image coordinates in ordered sequence in the horizontal lines.

i	24	27	61	65	68	69	75	84	98	100	100	101	108	110	138	140	149	186
j	59	51	125	115	132	104	133	89	150	72	77	158	79	89	107	190	105	135

Table 2 : Corner image coordinates ordered as a closed polygon (clockwise direction).

i	24	69	65	61	68	75	98	101	140	186	149	138	110	108	100	84	27	24
j	59	104	115	125	132	133	150	158	190	135	105	107	89	79	72	89	51	59

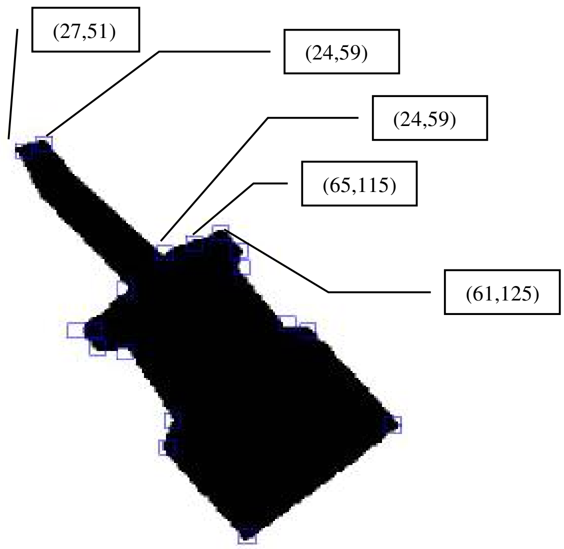


Figure 5: Object image and some corner coordinates

Figures (7) and (8) show the results of the calculated minimum errors using Eq. (4) obtained by comparisons between the geometry of the object models and the object images grabbed by the camera. With these experiments, it is possible to verify that the object observed in the camera comes closer to the geometry of the part (stored in the bank of models) as smaller is the minimum error found by Eq. (4). (Figs. (6-b), (6-d), (7-b) e (7-d)).

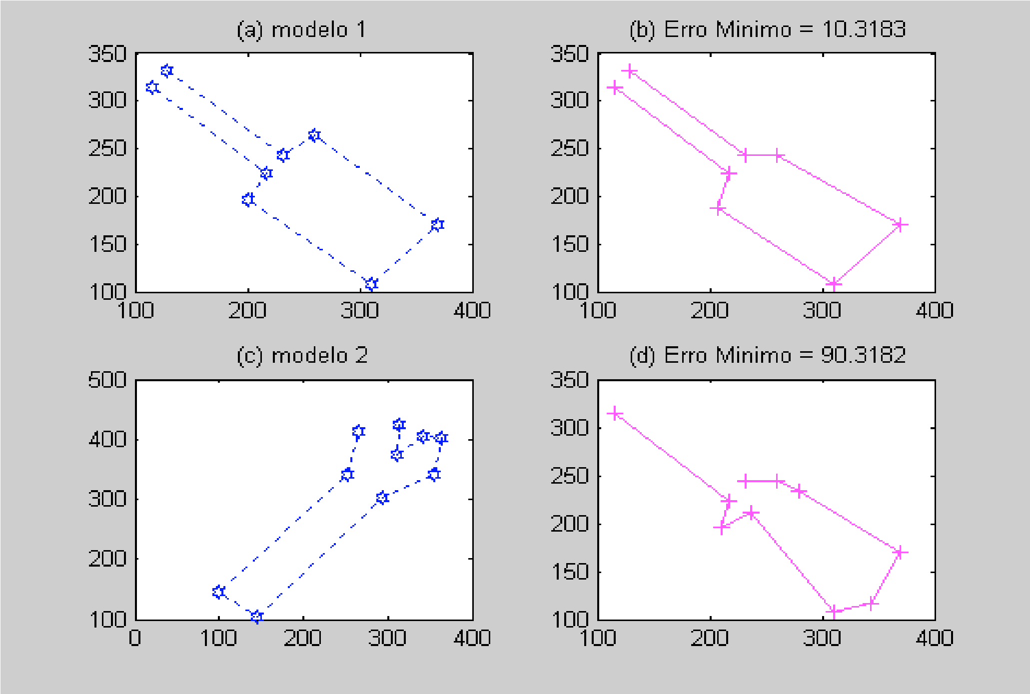


Figure 6. (a) Object model 1 geometry (b) corners in sequence compared with model 1 – Minimum error = 10.3183. (c) object model 2 geometry (d) corners in sequence compared to model 2 – Minimum error = 90.3182.

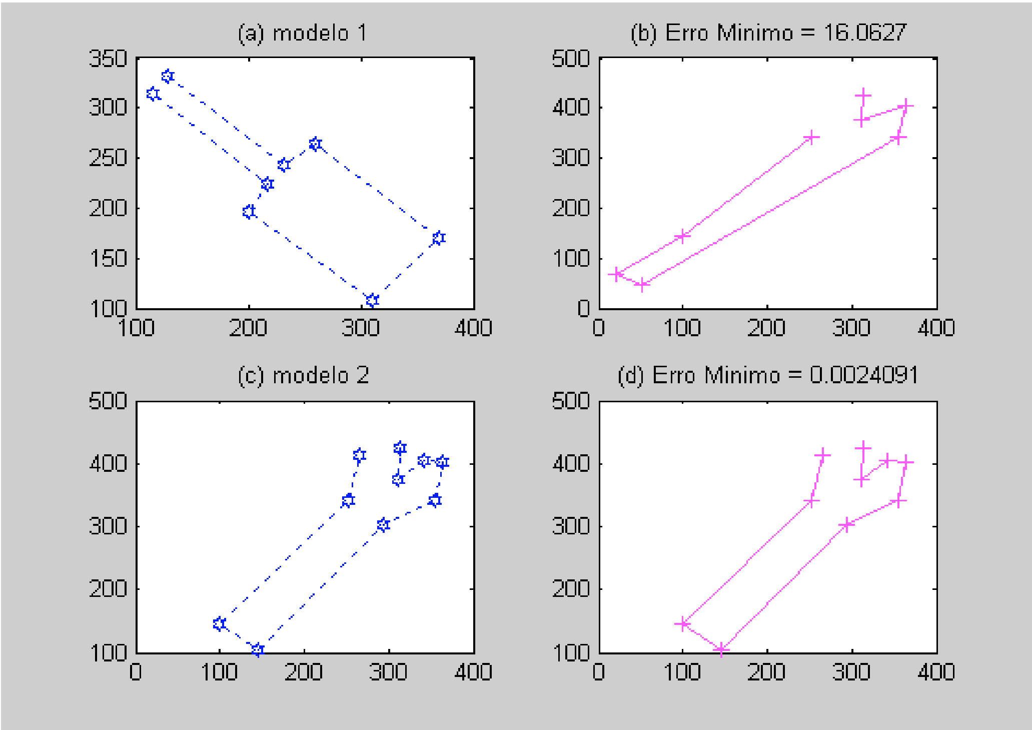


Figure 7. (a) Object model 1 geometry (b) corners in sequence compared with model 1 – Minimum error = 16.0627 (c) object model 2 geometry (d) corners in sequence compared to model 2 – Minimum error = 0.0024091

Figure (8) presents the object location results. Points C ', D ' and E ' had been determined by the solution of Eq. (7) and (8). The expected values are C '' = (4,5), D '' = (4,3) and E'' = (5,3), respectively.

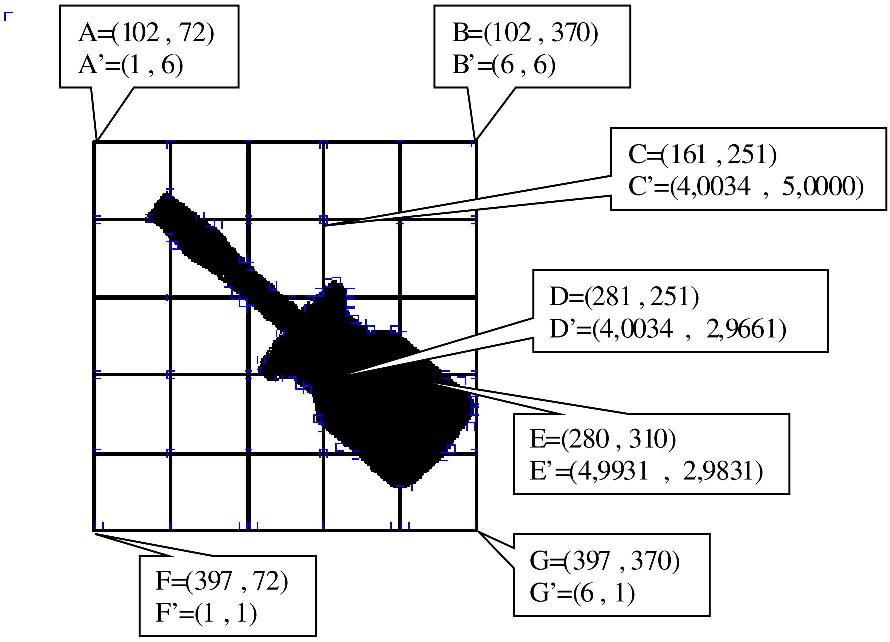


Figure 8. Object image with corners and its image coordinates and world coordinates – Location Results

The points represented by "`" are points in the 3-D space shown in Fig.(8) as: A ', B', C ', D ', E ', F ' and G '. The points, B, C, D, and, F and G are corners extracted from the image (pixel).

7. CONCLUSIONS

In this article the problem of object recognition was evaluated in images using a technique that uses geometric invariants defined with 5 coplanar points. Experimental results had shown the efficiency of the technique proposed. The creation of lists of corner points ordered in the clockwise direction and forming a closed polygon reduces considerably the amount of necessary combinations to compare selected object images to the model images stored as invariant vectors and consequently reducing time for processing.

Concerning the experiments to check the ability of the proposed system to locate objects identified in an image, it is possible to conclude that it is sufficient to know 4 non-collinear points in the 3-D space where the object was framed and its respective points in the camera image (pixels), to obtain the object position in camera coordinates. Therefore, the previous knowledge of the camera intrinsic and extrinsic parameters is not necessary.

Thus, since the objects are identified in an image and its location determined the manipulator can handle these objects with high degree of control of position and orientation, aiming at executing automatic tasks in industrial environments.

8. REFERENCES

Basak, J., 2000, "A Connectionist Model for Corner Detection in Binary and Gray Images", IEEE Transactions on Neural Networks, Vol. 11, no. 5, pp. 1124-1132.

Bong S. S., 2000, "A Target Recognition Technique Employing Geometric Invariants", Pattern Recognition, 33, pp. 413-425.

Hartley, R. and Zisserman, A., 2000, "Multiple View Geometry in Computer Vision", Cambridge University Press.

Horaud,R., Conio, B. and Leboulleux, O., 1989, "An Analytic Solution for the Perspective 4-Points problem", Computer Vision Graphics and Image Process, Volume 47, pp. 33-44.

Kuvich, G., 2005, "Scene understanding and objects detection and identification with a perceptual system based on the Network-Symbolic models for industrial robots", Proceedings of SPIE - The International Society for Optical Engineering, 5999.

Löpez-Franco, C., Bayro-Corrochano, E., 2006, Omnidirectional vision and invariant theory for robot navigation using conformal geometric algebra. Proceedings of the International Conference on Pattern Recognition. Vol. 1, Article number 1698957, Pages 570-573.

Lu, S, Wang, Z., Shen, J., 2003, "Neuro-fuzzy synergism to the intelligent system for edge detection and enhancement", Pattern and Recognition, Vol. 36, pp. 2395-2409.

Noble, J.A., 1988, "Finding Corners", Image and Vision Computer, V6, pp. 121-128.

Sumi, Y., Sallinen, M., Sirviö, M., Väinölä, J., 2004, "Recognition of large work objects in difficult industrial environments", Proceedings - IEEE International Conference on Systems, Man and Cybernetics, Volume 6, pp. 5285-5289 .

Trucco E., and Verri A., 1998, "Introductory Techniques for 3-D Computer Vision", Prentice Hall, USA, 343 p.

Vicente, M.A., Gil, P., Reinoso, O. and Torres, F., 2002, " Objects Recognition by Means of Projective Invariants Considering Corner-Points" , 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, ISSN 1213-6972, Volume 10, nº 1-2.

Zongmin Li; Jan Zhang; Yujie Liu; Hua Li, 2005, "The curve-structure invariant moments for shape analysis and recognition" , Computer Aided Design and Computer Graphics. Ninth International Conference on 7-10 Dec. 2005 Pages: 5 pp.

9. RESPONSIBILITY NOTICE

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