

Chapter Number

A Systems Engineering Approach to Robotic Bin Picking

Ovidiu Ghita and Paul F. Whelan
Dublin City University
Ireland

1. Introduction

In recent times the presence of vision and robotic systems in industry has become common place, but in spite of many achievements a large range of industrial tasks still remain unsolved due to the lack of flexibility of the vision systems when dealing with highly adaptive manufacturing environments. An important task found across a broad range of modern flexible manufacturing environments is the need to present parts to automated machinery from a supply bin. In order to carry out grasping and manipulation operations safely and efficiently we need to know the identity, location and spatial orientation of the objects that lie in an unstructured heap in a bin.

Historically, the bin picking problem was tackled using mechanical vibratory feeders where the vision feedback was unavailable. This solution has certain problems with parts jamming and more important they are highly dedicated. In this regard if a change in the manufacturing process is required, the changeover may include an extensive re-tooling and a total revision of the system control strategy (Kelley et al., 1982). Due to these disadvantages modern bin picking systems perform grasping and manipulation operations using vision feedback (Yoshimi & Allen, 1994).

Vision based robotic bin picking has been the subject of research since the introduction of the automated vision controlled processes in industry and a review of existing systems indicates that none of the proposed solutions were able to solve this classic vision problem in its generality. One of the main challenges facing such a bin picking system is its ability to deal with overlapping objects. The object recognition in cluttered scenes is the main objective of these systems and early approaches attempted to perform bin picking operations for similar objects that are jumbled together in an unstructured heap using no knowledge about the pose or geometry of the parts (Birk et al., 1981). While these assumptions may be acceptable for a restricted number of applications, in most practical cases a flexible system must deal with more than one type of object with a wide scale of shapes.

A flexible bin picking system has to address three difficult problems: scene interpretation, object recognition and pose estimation. Initial approaches to these tasks were based on modeling parts using the 2D surface representations. Typical 2D representations include invariant shape descriptors (Zisserman et al., 1994), algebraic curves (Tarel & Cooper, 2000),

conics (Bolles & Horaud, 1986; Forsyth et al., 1991) and appearance based models (Murase & Nayar, 1995; Ohba & Ikeuchi, 1997). These systems are generally better suited to planar object recognition and they are not able to deal with severe viewpoint distortions or objects with complex shapes/textures. Also the spatial orientation cannot be robustly estimated for objects with free-form contours. To address this limitation most bin picking systems attempt to recognize the scene objects and estimate their spatial orientation using the 3D information (Fan et al., 1989; Faugeras & Hebert, 1986). Notable approaches include the use of 3D local descriptors (Ansar & Daniilidis, 2003; Campbell & Flynn, 2001; Kim & Kak, 1991), polyhedra (Rothwell & Stern, 1996), generalized cylinders (Ponce et al., 1989; Zerroug & Nevatia, 1996), super-quadratics (Blane et al., 2000) and visual learning methods (Johnson & Hebert, 1999; Mittrapiyanuruk et al., 2004). The most difficult problem for 3D bin picking systems that are based on a structural description of the objects (local descriptors or 3D primitives) is the complex procedure required to perform the scene to model feature matching. This procedure is usually based on complex graph-searching techniques and is increasingly more difficult when dealing with object occlusions, a situation when the structural description of the scene objects is incomplete. Visual learning methods based on eigenimage analysis have been proposed as an alternative solution to address the object recognition and pose estimation for objects with complex appearances. In this regard, Johnson and Hebert (Johnson & Hebert, 1999) developed an object recognition scheme that is able to identify multiple 3D objects in scenes affected by clutter and occlusion. They proposed an eigenimage analysis approach that is applied to match surface points using the spin image representation. The main attraction of this approach resides in the use of spin images that are local surface descriptors; hence they can be easily identified in real scenes that contain clutter and occlusions. This approach returns accurate results but the pose estimation cannot be inferred, as the spin images are local descriptors and they are not robust to capture the object orientation. In general the pose sampling for visual learning methods is a problem difficult to solve as the numbers of views required to sample the full 6 degree of freedom for object pose is prohibitive. This issue was addressed in the paper by Edwards (Edwards, 1996) when he applied eigenimage analysis to a one-object scene and his approach was able to estimate the pose only in cases where the tilt angle was limited to 30 degrees with respect to the optical axis of the sensor.

In this chapter we describe the implementation of a vision sensor for robotic bin picking where we attempt to eliminate the main problem faced by the visual learning methods, namely the pose sampling problem. This paper is organized as follows. Section 2 outlines the overall system. Section 3 describes the implementation of the range sensor while Section 4 details the edge-based segmentation algorithm. Section 5 presents the viewpoint correction algorithm that is applied to align the detected object surfaces perpendicular on the optical axis of the sensor. Section 6 describes the object recognition algorithm. This is followed in Section 7 by an outline of the pose estimation algorithm. Section 8 presents a number of experimental results illustrating the benefits of the approach outlined in this chapter.

2. System Overview

The operation of the system described in this chapter can be summarized as follows (see Fig. 1). The range sensor determines the depth structure using two images captured with different focal settings. This is followed by the image segmentation process that decomposes

the input image into disjoint meaningful regions. The resulting scene regions from the image segmentation process are subjected to an orthographic projection that aligns them to be perpendicular on the optical axis of the sensor. This operation will determine 2 degrees of freedom (DOF) for each object (rotations about x and y axes). The recognition framework consists of matching the geometrical primitives derived from the segmented regions with those contained in a model database. The object that gives the best approximation with respect to the matching criteria is then referred to the pose estimation algorithm which constrain the object rotation around the optical axis of the range sensor (z axis) using a Principal Components Analysis (PCA) approach. Once the object pose is estimated, the grasping coordinates of the identified object are passed to the bin picking robot.

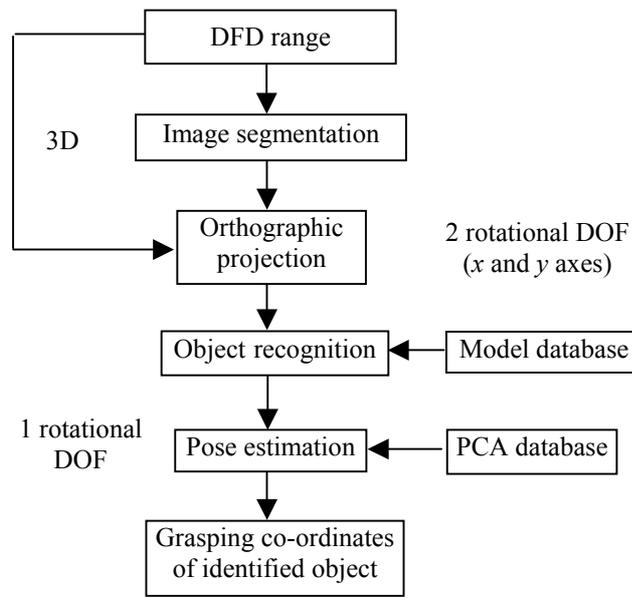


Fig. 1. Overall system architecture (Ghita & Whelan, 2003).

3. Range Sensor

The range sensor employed by this application is based on active depth from defocus (DFD). This ranging technique has been initially developed as a passive ranging strategy by Pentland (Pentland, 1987). The principle behind DFD range sensing extends from the fact that the scene objects are imaged in relation to their position in space. In this fashion, the objects that are placed on the focal plane are sharply imaged on the sensing element of the camera, while the points situated on the surface of the objects shifted from the focal plane are refracted by the lens into a patch whose size is in direct relationship with the distance from the focal plane to the imaged object. It has been demonstrated in (Subbarao, 1988; Nayar et al., 1995) that the diameter of the defocus (blur) patch is dependent on the object distance u , lens aperture D , sensor distance s and focal length f . While one image is not

sufficient to solve the uncertainty whether the scene object is placed in front or behind the focal plane, the depth can be uniquely estimated by measuring the blurring differences from two images captured with different focal settings. In our implementation the defocused images are captured by changing the sensor distance s (Ghita et al., 2005).

Since the level of blurriness in the image can be thought of as a convolution with a low pass filter (that is implemented by the point spread function (PSF)), to estimate the level of blurriness in the image we need to convolve the image with a focus operator that extracts the high frequency information derived from the scene objects (Pentland, 1987). Nonetheless this approach returns accurate results only if the scene objects are highly textured. When dealing with weakly and non-textured scene objects this approach returns imprecise depth estimation. To address this issue, a solution is to project a structured light onto the scene that forces an artificial texture on all visible surfaces of the scene. While the artificial texture has a known pattern, the focus operator is designed to respond strongly to the dominant frequency in the image that is associated with the illumination pattern (Girod & Scherrock, 1989; Nayar et al., 1995; Ghita et al., 2005).

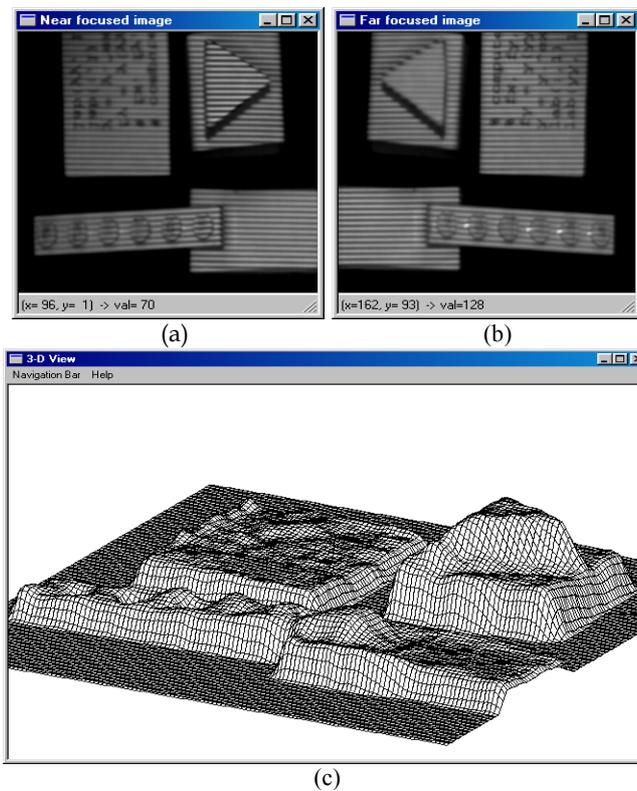


Fig. 2. Depth estimation for a scene defined by textureless, textured and mildly specular objects. (a) Near focused image. (b) Far focused image. (c) Depth estimation.

In our implementation we used an illumination pattern defined by evenly spaced opaque and transparent stripes and the focus operator is implemented by a tuned Gabor filter (full details about the implementation of our range sensor are provided in Ghita et al. 2005). Fig. 2 depicts the depth map obtained when the range sensor was applied to estimate the depth of a complex scene containing textureless, textured and specular objects.

4. Scene Segmentation Process

An important decision in developing robotic systems is to decide which sensorial information is better suited for a particular application. Henderson (Henderson, 1983) suggested to approach the scene segmentation using the information about the objects that define the scene. In this regard, if the scene objects are highly textured and depth discontinuities are significant best results will be achieved if range data is analysed. Conversely, if the scene is defined by small textureless objects better results may be obtained if the segmentation process is applied on intensity images (Ghita & Whelan, 2003).

While our application deals with the recognition of a set of textureless polyhedral objects we developed an edge-based segmentation scheme to identify the visible surfaces of the scene objects. Edges are associated with sharp transitions in pixel intensity distribution and they are extracted by calculating the partial derivatives in the input data. Edge detection is one of the most investigated topics in computer vision and to date there is no edge detector that is able to adapt to problems caused by image noise and low contrast between meaningful regions in the input data. Thus, the edge structure returned by the edge detector is either incomplete, gaps are caused by the low variation in the distribution of the input data, or contains false edges that are caused by image noise, shadows, etc. Thus after the application of edge detection, additional post-processing is applied to eliminate the spurious edge responses and bridge the gaps in the edge structure (this operation is also referred to as edge linking). Approaches that have been used to bridge the gaps in the edge structure include morphological methods (Hajjar & Chen, 1999), Hough transform (Davies, 1992), probabilistic relaxation techniques (Hancock & Kittler, 1990), multi-scale edge detection methods (Farag & Delp, 1995) and the inclusion of additional information such as colour (Saber et al., 1997). From these techniques the most common are the morphological and multi-scale edge linking strategies. In general, morphological edge linking techniques use the local information around edge terminators while multi-scale approaches attempt to bridge the gaps in the edge structure by aggregating the information contained in a stack of images with differing spatial resolutions. The main disadvantage associated with multi-scale approaches reside in the high computational cost required to calculate the image stack and in our implementation we developed a morphological edge linking scheme that evaluates the direction of edge terminators in identifying the optimal linking decisions.

4.1 Edge Linking

To extract the surfaces of the imaged scene objects we have developed a multi-step edge linking scheme that is used in conjunction with an edge detector that extracts the partial derivatives using the ISEF (Infinite Symmetrical Exponential Filter) functions (Shen & Castan, 1992). The reason to use the ISEF-based edge detector was motivated by the fact that its

performance in detecting true edges matches that achieved by the more ubiquitous Canny edge detector (Canny, 1986) but the computation of the ISEF edge detector entails a lower computational cost than that associated with the Canny edge detector. In our implementation we have set the scale parameter to 0.45 and the threshold parameters required by the hysteric threshold are selected using a scheme that minimise the incidence of small edge segments that are usually generated by image noise.

As mentioned earlier, the edge structure returned by the ISEF detector will be further post-processed using a multi-step morphological edge linking strategy. The first step of the edge linking algorithm (Ghita & Whelan, 2002) involves the extraction of the edge terminators (endpoints). The edgepoint extraction requires a simple morphological analysis where the edge structure is convolved with a set of 3×3 masks (Vernon, 1991). The second step of the algorithm determines the direction of the edge terminators by evaluating the linked edge points that generate the edge terminators. The application of the edge linking process for two iterations is illustrated in Fig. 3.

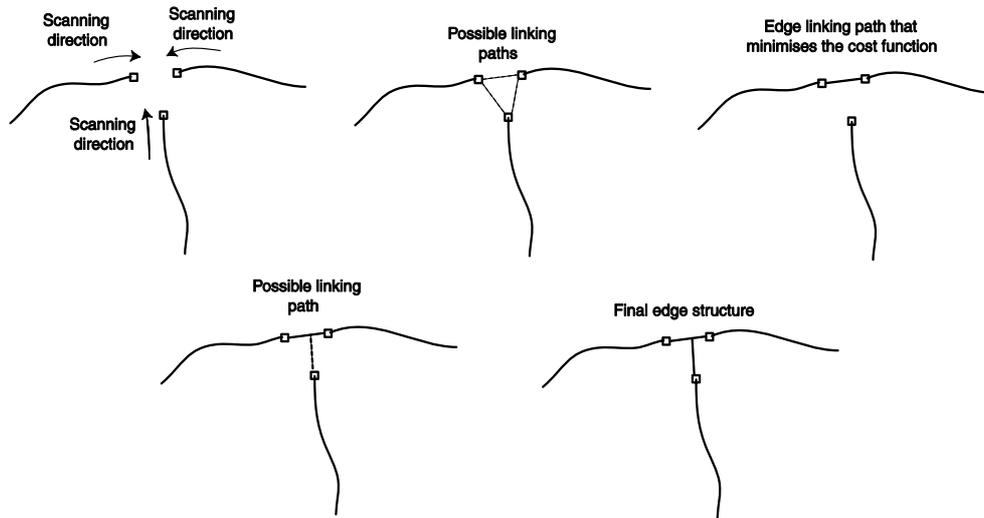


Fig. 3. The edge linking process. The algorithm evaluates all linking decisions around each edge terminator and the optimal linking path minimises the cost function depicted in equation (1). In this diagram the edge pixels are marked in black and the edge terminators are marked with a square box.

The third step of the edge linking scheme attempts to find the possible paths to bridge the gaps in the edge structure by analysing the edge pixels at the side indicated by the endpoint direction in an 11×11 neighbourhood. In this way, for each edge point situated in the endpoint's neighbourhood a linking factor is calculated using the following cost function,

$$Cost(ep) = k_d dist(et, ep) + k_e + k_{dir} \quad (1)$$

where et and ep are the co-ordinates of the endpoint and the edge pixel under analysis and $dist$ defines the Euclidean distance. In equation (1) k_d , k_{dir} and k_e are some pre-defined reward parameters (a detailed description of these parameters and a discussion in regard to their optimal selection is provided in Ghita & Whelan, 2002). The cost function is calculated for each edge pixel situated in the neighbourhood indicated by the endpoint direction and the minimal value determines the optimal linking path. The gap in the edge structure between the edge terminator and the edge pixel that returns the minimum linking factor is bridged using the Bresenham algorithm (Bresenham, 1965). Fig. 4 illustrates the performance of the edge linking algorithm when applied to an image detailing a cluttered scene.

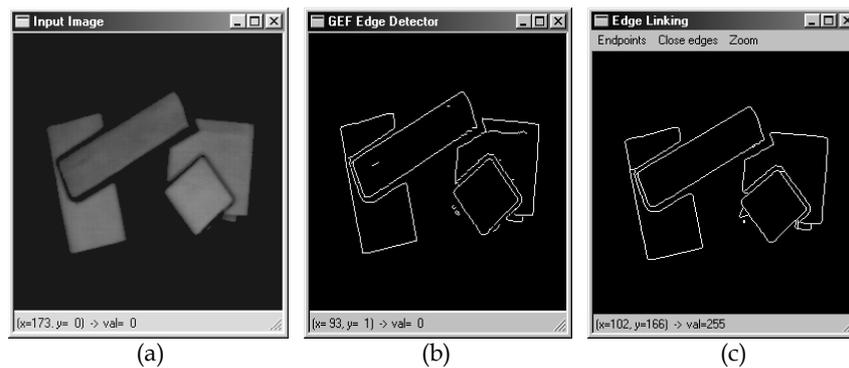


Fig. 4. The results of the scene segmentation process. (a) Input image. (b) Edge information. (c) Edge linking results. Note the removal of unconnected edge segments.

5. Data Formatting

Our application implements a vision sensor able to determine the information required by a bin picking robot to perform object manipulation. Since the objects of interest are polyhedral, a convenient representation is to describe them in terms of their surfaces that are identified by the scene segmentation algorithm detailed in the previous section. Thus, the object recognition task can be formulated in terms of matching the objects' visible surfaces with those stored in a model database. Although conceptually simple, this approach is quite difficult to be applied in practice since the geometrical characteristics of the object's surfaces are viewpoint dependent. To address this problem we need to align all visible surfaces resulting from the scene segmentation process to a planar that is perpendicular to the optical axis of the range sensor. In this fashion, we attempt to constrain two degrees of freedom (rotations about x and y axes) using the 3D information returned by the range sensor.

The first operation of the data formatting procedure involves the calculation of the normal vector for each surface resulting after the application of the scene segmentation procedure. Since the object surfaces are planar, the normal vector can be calculated using the knowledge that elevation (z co-ordinate) is functionally dependent on the x and y co-

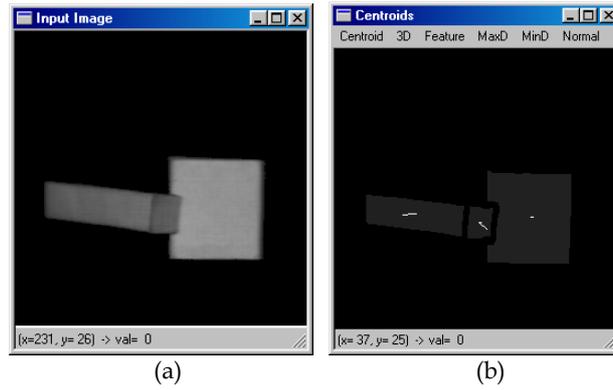
ordinates (i.e. $z = a_1x + a_2y + a_3$). Then given a set of n points from range data that belong to the segmented surface, the normal vector can be statistically computed by a planar fitting of the 3D points as follows,

$$Err(\hat{a}) = \sum_{i=1}^n (\hat{a}_1x_i + \hat{a}_2y_i + \hat{a}_3 - z_i)^2 \quad (2)$$

where $\hat{a} = [\hat{a}_1, \hat{a}_2, \hat{a}_3]$ are the estimated values. Equation (2) generates a simultaneous system where the unknown values are \hat{a} . The normal vector associated with the surface under analysis is represented in homogenous form as $\bar{N} = [n_x, n_y, n_z, 1]^T = [\hat{a}_1, \hat{a}_2, -1, 1]^T$ (Ghita et al., 2007). As mentioned previously, our aim is to calculate the rotations about x and y axes. The rotation angle about x axis A_x is calculated using the following expression: $A_x = \tan^{-1}(n_y/n_z)$. The rotation angle about y axis is computed using the transform $N_{Rx} = R_x N = [n_{rx}, n_{ry}, n_{rz}, 1]^T$, as $A_y = -\tan^{-1}(n_{rx}/n_{rz})$, where \tan^{-1} is the four quadrant inverse tangent. Once the angles A_x and A_y are estimated, the required transformation that is applied to align the surface under analysis to the planar perpendicular to the axis of the range sensor can be formulated as follows,

$$H = T_0^{-1} R_y R_x T_0 \quad (3)$$

where T_0 is the transformation that translates the 3D points that define the surface about the origin and R_x and R_y are the rotation matrices about x and y axes. Fig. 5 illustrates the results obtained after the application of the orthographic projection.



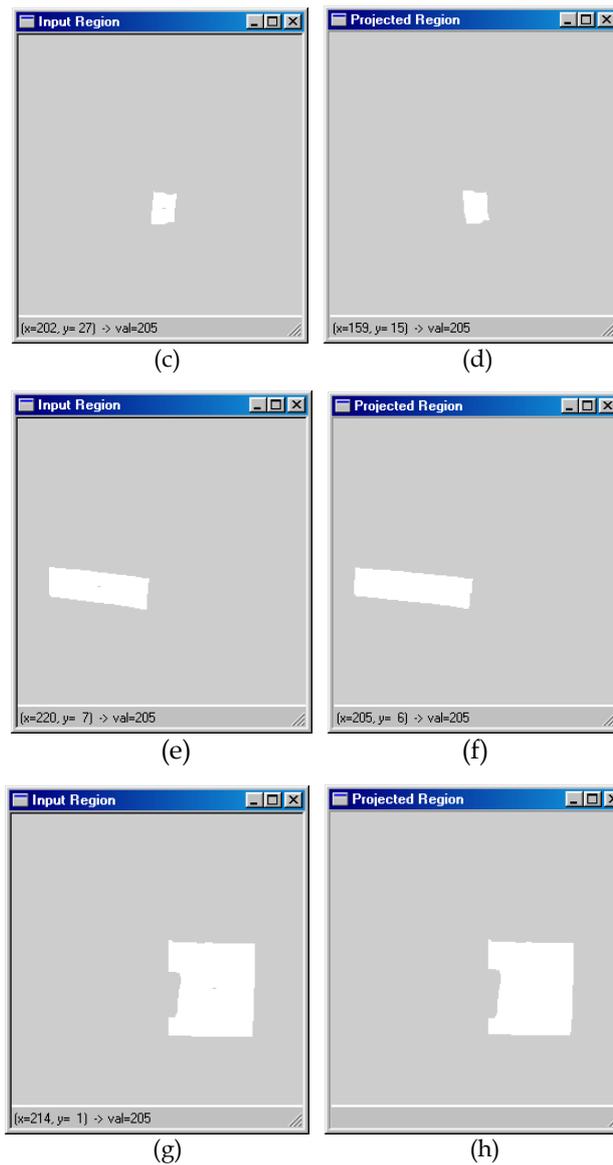


Fig. 5. Orthographic projection of the segmented scene regions. (a-b) Input image and scene regions resulting from the segmentation process (normal vectors relative to the range sensor position). (c-d) Orthographic projection of the first region ($A_x = 26.79^\circ$, $A_y = -18.61^\circ$). (e-f) Orthographic projection of the second region ($A_x = -36.19^\circ$, $A_y = -4.76^\circ$). (g-h) Orthographic projection of the third region ($A_x = 6.08^\circ$, $A_y = 2.68^\circ$).

6. Object recognition

As indicated in the previous section, the recognition of scene objects is formulated as the recognition of their visible surfaces resulting after the application of the scene segmentation process using an approach that calculates features that sample the geometrical properties of the object surfaces. While the geometric characteristics of the object surfaces are dependent on their orientation in space, in order to eliminate the viewpoint distortions the segmented surfaces were subjected to a 3D data formatting procedure that aligns them to a planar whose normal vector is aligned to the optical axis of the range sensor (z axis). The next step of the algorithm deals with the extraction of geometrical primitives that are used to perform the scene to model recognition process. Approaches that have been used include the extraction of local features such as junctions, lines and partial contours (Bolles & Horaud, 1986; Lowe, 2004) and macro features such as area, perimeter and statistical features (Ghita & Whelan, 2003). Local features may appear better suited when dealing with scenes affected by clutter and occlusions than macro features. But it is useful to note that approaches based on local features rely on a detailed structural description of the objects of interest and when dealing with complex scenes a large number of hypothesis are generated, a fact that requires the development of complex scene to model matching procedures. While our goal is the recognition of a set of polyhedral objects, macro features represent a better option since the segmented surfaces are planar and they can be easily indexed to describe the object structure. To this end, we have adopted features such as area, perimeter, shape factor and radii (maximum and minimum) distances calculated from the surface's centroid to the surface border (Ghita & Whelan, 2003).

The developed object recognition algorithm consists of two main stages. The training stage consists of building the database by extracting the aforementioned features for each surface of the object. Since the features involved have different ranges, to compensate for this issue we have applied a feature normalisation procedure where each feature is normalised to zero mean and unit variance (Duda et al., 2001). The matching stage consists of computing the Euclidean distance between the normalised features calculated for scene surfaces and object surfaces contained in the model database.

$$dist_j = \sqrt{\sum_{i=1}^n (X_j[i] - Y[i])^2} \quad \text{for } i = 1, \dots, n \quad (4)$$

where X_j is the j^{th} pattern contained in the model database and Y defines the pattern derived from an input region. The input scene surface is contained in the database if the minimum distance that gives the best approximation is smaller than a predefined threshold value.

One problem with this approach is the fact that most scene surfaces are affected by occlusions. As the object recognition algorithm is included in the development of a robotic application, we focus the attentions only on the topmost objects since they can be easily manipulated and their surfaces are not affected by severe occlusions. The selection of the topmost object is achieved by eliminating the surfaces that are affected by occlusions based on the 3D information supplied by the range sensor. The scene to model verification procedure is applied only for surfaces that pass the 3D selection criteria (for additional details refer to Ghita & Whelan, 2003).

7. 3 DOF Pose Estimation

The orthographic transformation illustrated in equation (3) can constrain only two degrees of freedom (DOF), the rotations about x and y axes. The surfaces subjected to this orthographic transformation are perpendicular on the axis of the range sensor and the estimation of the surface rotation about z axis can be carried out using Principal Components Analysis (PCA). This procedure involves the calculation of an eigenspace representation from a set of training images that are generated by rotating the object surfaces in small increments. To estimate the rotation about z axis, all recognized scene surfaces are projected onto the eigenspace and their projections are compared to those stored in the model database (whose rotations about the z axis are known). The minimal distance between the projection of the input surface and those contained in the model database gives the best match.

8. Experiments and Results

The vision sensor detailed in this chapter consists of four main components, range sensing, scene segmentation, object recognition and pose estimation. Our implementation employs an active DFD range sensor whose implementation has been outlined in Section 3. To test the performance of the developed range sensor we have applied it to recover the depth information from scenes defined by textured and textureless objects. The relative accuracy was estimated for successive measurements and was formulated as the maximum error between the real and estimated depth values. During the operation the range sensor was placed at a distance of 86cm above the baseline of the workspace. The relative accuracy attained by the developed sensor when applied to scenes containing non-specular objects with bright surfaces is 3.4% normalised in agreement with the distance from the sensor.

The developed bin picking system it has been applied to 5 different polyhedral objects that are used to create various cluttered scenes. The edge-based segmentation algorithm detailed in Section 4 is applied to identify the object surfaces. The surfaces resulting after the application of the scene segmentation algorithm are subjected to data formatting in order to constrain 2 rotational DOF (rotations about x and y axes). Since data formatting involves 3D analysis, the precision of this procedure is influenced by the accuracy of the depth estimation. The performance of the data formatting procedure is illustrated in Figs. 6 and 7.

The third major component of the algorithm addresses the object recognition task. The algorithm was able to identify the topmost objects in all situations and is able to identify correctly the scene objects if the occlusion cover less than 20% of the object's total surface. The last component of the algorithm is applied to identify the rotation about z axis. In our implementation we have created a PCA model database for each object of interest and the object rotation has been sampled uniformly by acquiring 24 training images with the object lying flat on a dark worktable. This generates 24 PCA projections that are able to sample the object rotation with a resolution of 15 degrees. To increase the resolution of the PCA projections we have applied a linear interpolation procedure that generate 30 interpolated projections between any adjacent projections generated by the 24 images contained in the training set (Ghita & Whelan, 2003; Ghita et al., 2007). The performance of the pose estimation is affected by the accuracy of the data formatting procedure and the experiments indicate that the pose is more precise for low values of the tilt angles (rotations about x and y

axes). This is motivated by the relative low resolution of the range sensor in sampling depth discontinuities. In our experiments the rotation about z axis was measured with an error of 2.1 degree under the condition that the rotations about x and y axes are smaller than 25 degrees.

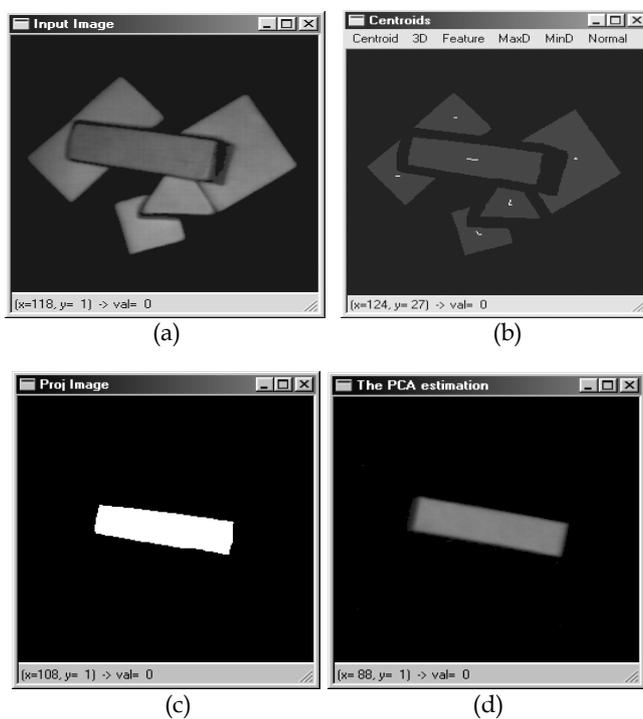


Fig. 6. Complex object scene affected by clutter and occlusions (Ghita et al., 2007). (a) Input image. (b) Surface segmentation. (c) Orthographic projection for best estimated surface ($A_x = -20.94^\circ$, $A_y = 4.21^\circ$). (d) PCA estimation.

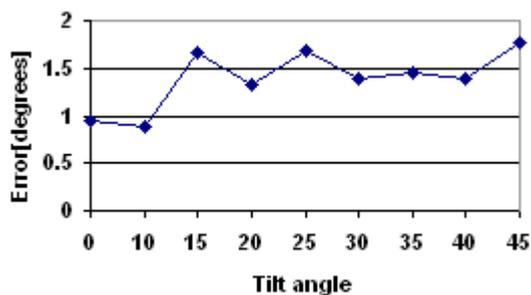


Fig. 7. Data formatting estimation accuracy (rotation about x axis).

9. Conclusions

This chapter describes the development of a fully integrated vision sensor for robotic bin picking. The developed vision sensor is able to provide the information to a bin picking robot to perform scene understanding and object grasping/manipulation operations. Our implementation employs a range sensor based on active depth from defocus that is used in conjunction with a multi-stage scene understanding algorithm that is able to identify and estimate the 3D attitude of the scene objects. In this regard, the scene segmentation scheme attempts to separate the scene regions that are associated with object surfaces using an edge based implementation. The novel part of this scheme is the edge linking procedure that is able to return quality connected edge structures. The object recognition scheme performs scene to model verification using the global attributes extracted from the segmented scene surfaces. As these features are vulnerable to viewpoint distortions we have devised a data formatting scheme that re-format the orientation of the scene surfaces on a planar perpendicular on the optical axis of the sensor. This transformation eliminates the viewpoint distortions and allows us to apply standard PCA to sample the rotation about z axis. The experimental results indicate that reasonable accurate pose estimation is obtainable from this approach and we believe that the developed vision sensor is particularly useful when applied to scenes defined by polyhedral objects or objects with well-defined surfaces.

10. References

- Ansar, A. & Daniilidis, K. (2003). Linear pose estimation from points or lines, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 578-589.
- Birk, J.R.; Kelley, B. & Martins, H. (1981). An orienting robot for feeding workpieces stored in bins, *IEEE Trans. Syst. Man Cybern.*, vol. 11, no. 2, pp. 151-160.
- Blane, M.; Lei, Z.; Civi, H. & Cooper, D.B. (2000). The 3L algorithm for fitting implicit polynomial curves and surfaces to data, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no.3, pp. 298-313.
- Bolles, R.C. & Horaud P. (1986). 3DPO: A three dimensional part orientation system, *Intl. J. Robotics Res.*, vol. 5, no. 3, pp. 3-26.
- Bresenham, J.E. (1965). Algorithm for computer control of a digital plotter, *IBM Systems Journal*, vol. 4, no. 1, pp. 25-30.
- Campbell, R. & Flynn, P. (2001). A survey of free-form object representation and recognition techniques, *Computer Vision and Image Understanding*, vol. 81, no. 2, pp. 166-210.
- Canny, J. (1986). A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 698-700.
- Davies, E.R. (1992). Locating objects from their point features using an optimised Hough-like accumulation technique, *Pattern Recognition Letters*, vol. 13, no. 2, pp. 113-121.
- Duda, R.O.; Hart, P.E. & Stork, D.G. (2001). *Pattern classification*, (2nd edition), Wiley, ISBN 0-471-05669-3, USA.
- Edwards, J. (1996). An active, appearance-based approach to the pose estimation of complex objects, *Proc. of the IEEE Intelligent Robots and Systems Conference*, Osaka, Japan, pp. 1458-1465.

- Fan, T.; Medioni, G. & Nevatia, A. (1989). Recognizing 3-D objects using surface description, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 11, pp. 1140-1157.
- Farag, A. & Delp, E.J. (1995). Edge linking by sequential search, *Pattern Recognition*, vol. 28, no. 5, pp. 611-633.
- Faugeras, O.D. & Hebert, M. (1986). The representation, recognition and locating of 3-D objects, *Intl. J. Robotics Res.*, vol. 5, no. 3, pp. 27-52.
- Forsyth, D.; Mundy, J.L.; Zisserman, A.; Coelho, C.; Heller, A. & Rothwell C. (1991) Invariant descriptors for 3-D object recognition and pose, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 10, pp. 971-991.
- Ghita, O. & Whelan, P.F. (2002). Computational approach for edge linking, *Journal of Electronic Imaging*, vol. 11, no. 4, 2002, pp. 479-485.
- Ghita O. & Whelan, P.F. (2003). A bin picking system based on depth from defocus, *Machine Vision and Applications*, vol. 13, no. 4, pp. 234-244.
- Ghita, O; Whelan, P.F. & Mallon, J. (2005). Computational approach for depth from defocus, *Journal of Electronic Imaging*, vol. 14, no. 2, 023021.
- Ghita, O.; Whelan, P.F.; Vernon D. & Mallon J. (2007). Pose estimation for objects with planar surfaces using eigenimage and range data analysis, *Machine Vision and Applications*, vol. 18, no. 6, pp. 355-365.
- Girod, B. & Scherrock, S. (1989). Depth from defocus and structured light, *Optics, Illumination and Image Sensing for Machine Vision IV, Proc. of the Soc. of Photo-Opt. Instrum.*, vol. 1194, pp. 209-215.
- Hajjar, A. & Chen, T. (1999). A VLSI architecture for real-time edge linking, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 89-94.
- Hancock, E.R. & Kittler, J. (1990). Edge labelling using dictionary-based relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 2, pp.165-181.
- Henderson, C. (1983). Efficient 3-D object representation for industrial vision systems, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, no. 6, pp. 609-617.
- Johnson, A. & Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3D scenes, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433-449.
- Kelley, B.; Birk, J.R.; Martins, H. & Tella R. (1982). A robot system which acquires cylindrical workpieces from bins, *IEEE Trans. Syst. Man Cybern.*, vol. 12, no. 2, pp. 204-213.
- Kim, W. & Kak, A. (1991). 3-D object recognition using bipartite matching embedded in discrete relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 224-251.
- Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints, *Intl. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110.
- Mittrapiyanuruk, P.; DeSouza, G.N. & Kak, A. (2004). Calculating the 3D-pose of rigid objects using active appearance models, *Intl. Conference in Robotics and Automation*, New Orleans, USA.
- Murase, H. & Nayar, S.K. (1995). Visual learning and recognition of 3-D objects from appearance, *Intl. Journal of Computer Vision*, vol. 14, pp. 5-24.
- Nayar, S.K.; Watanabe, M. & Noguchi, M. (1995). Real-time focus range sensor, *Proc. of the Intl. Conf. on Computer Vision (ICCV)*, pp. 995-1001.
- Ohba, K. & Ikeuchi, K. (1997). Detectability, uniqueness and reliability of eigen window for stable verification of partially occluded objects, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 1043-1048.

- Pentland, A. (1987). A new sense for depth of field, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, pp. 523-531.
- Ponce, J.; Chelberg, D. & Mann, W. (1989). Invariant properties of straight homogenous generalized cylinders and their contours, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 9, pp. 951-966.
- Rothwell, C. & Stern, J. (1996). Understanding the shape properties of trihedral polyhedra, *Proc. of European Conference on Computer Vision*, 1996, pp. 175-185.
- Saber, E.; Tekalp, A.M. & Bozdagi, G. (1997). Fusion of color and edge information for improved segmentation and edge linking, *Image and Vision Computing*, vol. 15, no. 10, pp. 769-780.
- Shen, J. & Castan, S. (1992). An optimal operator for step edge detection, *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 2, pp. 112-133.
- Subbarao, M. (1988). Parallel depth recovery by changing camera parameters, *Proc. of the IEEE Conf. on Computer Vision*, pp. 149-155.
- Tarel, J.P. & Cooper, D.B. (2000). The complex representation of algebraic curves and its simple exploitation for pose estimation and invariant recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 663-674.
- Vernon, D. (1991). *Machine vision: Automated visual inspection and robot vision*, Prentice-Hall International, ISBN 0-13-543398-3, UK.
- Yoshimi, B.H. & Allen, P. (1994). Visual control of grasping and manipulation tasks, *Proc. of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Las Vegas, USA.
- Zerroug, M. & Nevatia, R. (1996). 3-D description based on the analysis of the invariant and cvasi-invariant properties of some curved-axis generalized cylinders, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 237-253.
- Zisserman, A.; Forsyth, D.; Mundy, J.; Rothwell, C.; Liu, J. & Pillow, N. (1994). 3D object recognition using invariance, Technical report, Robotics Research Group, University of Oxford, UK.