Research article

Laser ranging and video imaging for bin picking

Faysal Boughorbel
Yan Zhang
Sangkyu Kang
Umayal Chidambaram
Besma Abidi
Andreas Koschan and
Mongi Abidi

The authors

Faysal Boughorbel, Yan Zhang, Sangkyu Kang, Umayal Chidambaram, Besma Abidi, Andreas Koschan and **Mongi Abidi** are all based at Imaging, Robotics and Intelligent Systems Lab, Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville, USA.

Keywords

Machine vision, Order picking, Imaging, 3D, Modelling

Abstract

This paper describes an imaging system that was developed to aid industrial bin picking tasks. The purpose of this system was to provide accurate 3D models of parts and objects in the bin, so that precise grasping operations could be performed. The technology described here is based on two types of sensors: range mapping scanners and video cameras. The geometry of bin contents was reconstructed from range maps and modeled using superquadric representations, providing location and parts surface information that can be employed to guide the robotic arm. Texture was also provided by the video streams and applied to the recovered models. The system is expected to improve the accuracy and efficiency of bin sorting and represents a step toward full automation.

Assembly Automation
Volume 23 · Number 1 · 2003 · pp. 53–59
© MCB UP Limited · ISSN 0144-5154
DOI 10.1108/01445150310460097

1. Introduction

An important quest in the field of production line automation is the design of flexible and autonomous robotic systems that can grasp and manipulate complex objects. Most current systems depend on complete knowledge of both the shape and position of the parts. Therefore, each time there is a need to handle a new series, the line must be adapted extensively. These changes can be very expensive, especially for small series. Hence, there is a need for efficient manipulators that can recognize complex and unorganized objects with little or no prior knowledge about the pose and geometry of the parts. Extensive research went into the design of vision systems to help achieve this goal. The task is made more difficult, given the industrial standard requirements of precision, robustness, and speed.

Kak and Edwards (1995) summarized their work on bin picking, focusing on feature-based methods and geometric hashing techniques. They concluded that the state of the art allows for industrial applications in the case of simple objects such as cylinders. To recognize and grasp objects Rahardja and Kosaka (1996) based their work on stereo images and extracted features such as polygonal and circular parts from complex objects. The system incorporated object identification and pose estimation in a single step. Ikeuchi (1987) used CAD models of objects and built representation trees to determine optimal features for recognition and grasping. Several other researchers used range images, along with various segmentation and recognition techniques (Al-Hujazi and Sood, 1990).

While most vision systems that are applied to bin picking rely on either video or range imagery, we present in this paper an approach that combines both modalities. The use of laser range maps will provide accurate bin geometry aiding grasp planning. Geometric models augmented with color information from the video input will help both human operators, as well as automatic object extraction modules, in segmenting and extracting parts of interest. In addition, video

This work was supported by the University Research Program in Robotics under grant DOE-DE-FG02-86NE37968, and by the DOD/TACOM/NAC/ARC Program, R01-1344-18, by FAA/NSSA Program, R01-1344-48/49.

streams are utilized for robot arm tracking, which increases the robustness of the overall system. We also describe the use of superquadrics-based object segmentation and recognition to increase the precision of parts grasping. Among the applications that our research targets is the task of sorting nuclear waste at US Department of Energy facilities. In this context we are contributing to the design of imaging and vision components of a radioactive mixed waste sorting station. In the next section, we will present an overview of the proposed system.

2. System overview

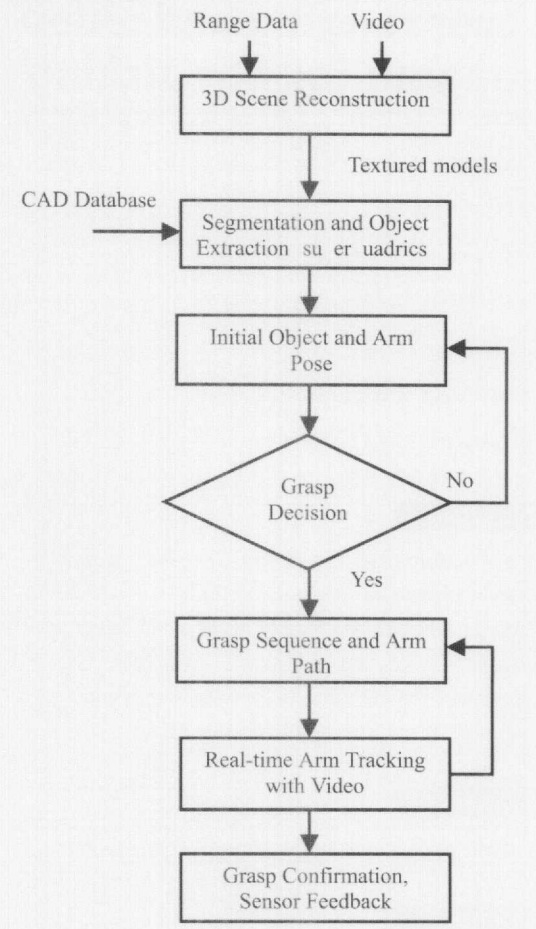
At the core of the proposed system for bin picking are several computer vision techniques that are employed for scene reconstruction and for object tracking in video sequences. The primary sensors used in this work are:

(1) a high-resolution laser range scanner that will acquire the geometry of the scene, and
(2) a calibrated color video camera.

The first module in the pipeline shown in Figure 1 deals with 3D reconstruction. At this level, 3D polygon models are recovered from the range images. We assume fixed sensor geometry, where the relative position of the video camera and of the scanner is known. This will allow for accurate range-video registration and therefore enable texture mapping of the reconstructed 3D bin model with color images. The augmentation of geometry information with color can be very useful for object segmentation and recognition, which is the second step in the system. If some *a priori* knowledge about the objects in the bin is available in the form of CAD models, the task of isolating these objects is greatly simplified. In the case of parts of unknown geometry, another object representation technique, known as superquadrics, will be used.

The proposed system can handle both cases of bins containing either mostly identical shaped objects or objects with different and unknown shapes. When objects of the same shapes (geometries) are present in a bin, superquadrics can classify the objects successfully because superquadrics not only provide quantitative shape information about the objects, but also information about their position and orientation. So basically, there is a one-to-one mapping between objects located in the bin and corresponding recovered superquadrics. Similar or same shaped objects contained in a bin will not add extra difficulties to the bin picking problem in terms of superquadric representation. As for the previously mentioned task of sorting mixed nuclear waste, there exist no information about the shapes of the objects that were randomly put into bins.

Figure 1 Flowchart of the proposed vision guided grasping system



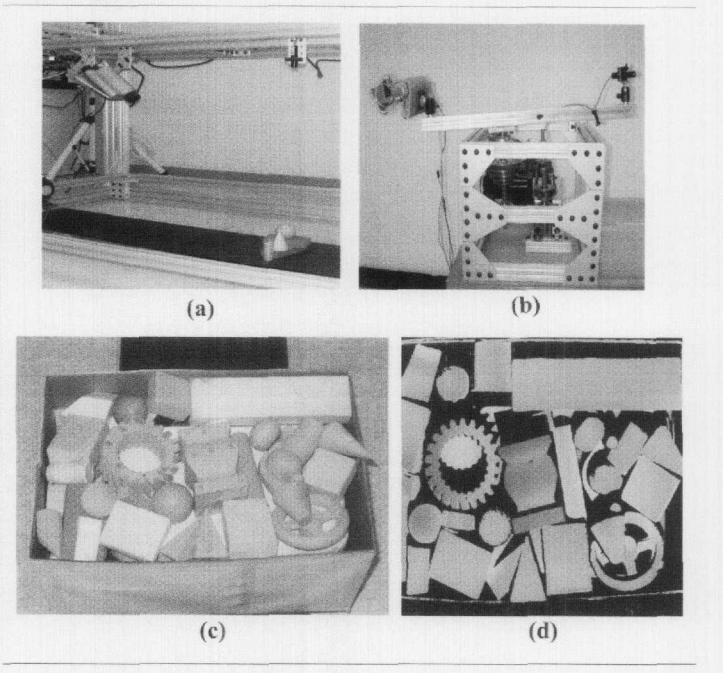
After segmenting and representing the part of interest and determining its position, a decision module checks object surfaces for possible grasp configurations. It also checks for the interference of other bin parts with the grasping process. If the decision is a "Go", a grasping sequence is computed and implemented. Otherwise, other elements of the bin will be examined. During the grasping sequence, a real-time feedback on the position of the robot arm is needed in order to adjust the motion and correct the possible errors. This is performed through robust video tracking techniques and real-time pose

Assembly Automation
Volume 23 · Number 1 · 2003 · 53–59

estimation. Finally, the hand-object contact is confirmed when sensory input from both the vision system and touch sensors allows for a positive decision. For the successful operation of the vision-guided grasping system, emphasis was put on the accuracy of the bin reconstruction as well as on the precision of the pose estimation module. The system therefore includes a closed control loop that minimizes the effect of error drift in different sub-modules.

In the following sections we will present in more detail the different tasks that were addressed toward a fully operational system. We start Section 3 with scene reconstruction, where we describe our approach to the recovery of geometrically accurate overlaid 3D models from range and color data. In Section 4 we show a technique for segmenting, recognizing, and representing bin objects based on superquadrics. Section 5 will present our approach to the validation of the robot arm position during the grasping process.

Figure 2 Laser range scanner used in our experiments, mounted in a linear (a) and rotational (b) scanning configuration. Bin objects (c) and the acquired range image (d)



3. 3D reconstruction of bin objects

To build a 3D model of bin objects, it is necessary to use range acquisition devices to map the geometry of the scene. Two approaches are commonly used to perform this task: passive and active range sensing methods. The first approach relies primarily on stereovision, which infers depth by computing the disparity between two images. While stereo has several advantages such as real-time implementation and relative low cost, some still unresolved problems such as matching ambiguities reduce the accuracy and reliability of the resulting depth maps. Nevertheless, stereo methods are commonly used in robotic applications. The second approach relies mostly on laser range sensing. In this case time of flight and triangulation methods were implemented. Recently, laser range scanners have become increasingly affordable and fast. In our experiments we used an accurate (sub-millimeter) triangulation-based range profiling system (Figure 2).

A range map acquired from a given point of view provides a good estimate of the scene structure. However, single view scans are not sufficient to describe the objects fully, because of occlusion problems. This is why range images do not provide complete 3D information and are known as $2\frac{1}{2}$D data. For these reasons, multi-view 3D registration is an important scene reconstruction step. Several approaches were proposed in the literature to perform this task. Some methods use feature extraction and description to perform 3D matching (Campbell and Flynn, 2001). The goal of these techniques is to design invariant feature descriptors that are robust to noise. Such methods are mostly used in the preliminary stage of 3D registration. For fine registration the predominant method is the so-called iterative closest point (ICP) algorithm (Besl and McKay, 1992) and its modifications, which minimize geometric distances between the datasets. Once the different views were registered, an integration step is performed to extract scene surfaces and to obtain a full 3D model. The quality of the model depends on the number of acquired scans and the viewpoint in the scene.

In addition to recover the shape of bin objects, we also want to use color information available from video and high-resolution still cameras. These images will be used as a texture map for the recovered scene models. Here again sensor registration is the first step that needs to be performed. Registering the cameras with the scanner is required only once if the relative position of the imaging devices is fixed. The task is very similar to camera calibration (Tsai, 1987), requiring the

establishment of a set of 3D to 2D correspondences and the computation of the internal and external parameters of the color camera. Given the camera matrix, computer graphics techniques are used to overlay texture on top of the 3D model (Figure 3). Besides color images, other imaging modalities could be used in addition to range data.

4. Superquadrics-based 3D object representation for bin picking

In addition to object modeling, some of the most important tasks for automated bin picking are object recognition, path planning, and grasp point determination. To increase the accuracy of these tasks we used superquadric representations, which provide precise models of bin objects. This approach also significantly reduces the amount of geometric information that needs to be processed since it describes the objects using just a few parameters.

Volumetric primitives, as the highest-level representation, represent the most intuitive decomposition of an object into parts. Models composed of volumetric primitives can easily support part articulation and, at the structural level, are insensitive to dimensional changes in the parts. These part-level characteristics enable volumetric primitives to support object manipulation, functional-based object recognition, and other high-level activities. Several volumetric primitives for modeling object parts have been used in computer visions. The most commonly used are generalized cylinders, geons, superquadrics, deformable superquadrics, and many other volumetric primitives. These part-level models can be classified into two categories: qualitative (non-parametric) and quantitative (parametric) models.

Superquadrics, as parametric models, appeared in computer vision as an answer to some of the problems with generalized cylinders. As a subclass of generalized cylinders, superquadrics are a family of geometric solids, which can be interpreted as a generalization of basic quadric surfaces and solids. With only a few parameters, they can represent a large variety of common geometric solids as well as smooth shapes (Figure 4). This makes superquadrics much more convenient for object representation. Moreover, superquadrics can be deformed globally and locally by stretching, bending, tapering or twisting, and then combined using

Figure 3 An overview of the system for 3D modeling from range and color images

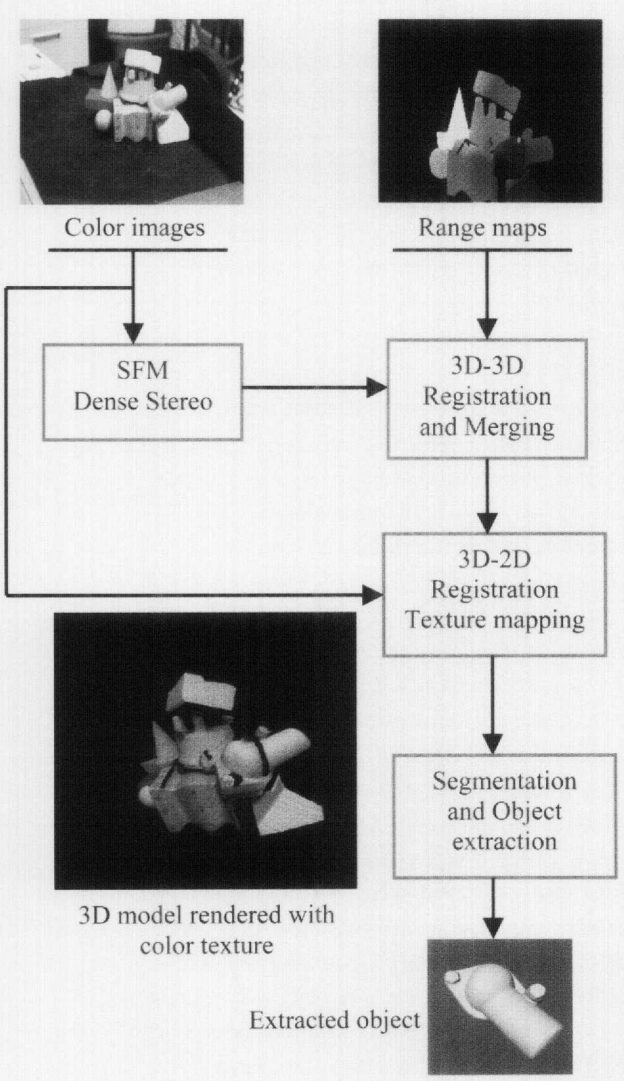
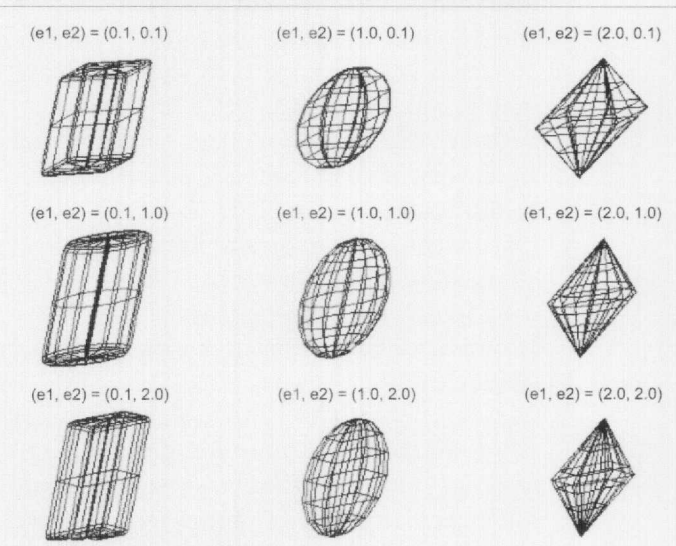


Figure 4 Superquadric shapes with various shape parameters



Boolean operations to build more complicated objects. One of the most attractive features of superquadrics is their interchangeable implicit and explicit defining function. The implicit definition is differentiable everywhere and is especially suitable for model recovery. On the other hand, the explicit form is quite convenient for visual rendering. Objects represented by superquadric models can be recognized easily in object recognition tasks.

Components of a typical object recognition system are shown in Figure 5. Primitive extraction is the first step in a recognition task and the performance of the recognition system significantly depends on the primitives used. Among all the primitives, volumetric primitives provide a powerful indexing mechanism since they represent an intuitive decomposition of objects into parts.

The implicit definition of superquadrics is expressed (Barr, 1984) as

$$F(x,y,z) = \left[\left(\frac{x}{a_1}\right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y}{a_2}\right)^{\frac{2}{\varepsilon_2}}\right]^{\frac{\varepsilon_1}{\varepsilon_2}} + \left(\frac{z}{a_3}\right)^{\frac{2}{\varepsilon_1}} = 1, \quad \varepsilon_1, \varepsilon_2 \in (0,2) \tag{1}$$

where (x, y, z) are the coordinates of a point of the model, (a_1, a_2, a_3) the superquadric sizes in the (x, y, z) directions, and $(\varepsilon_1, \varepsilon_2)$ the shape factors. To describe a superquadric model in the world coordinate system, 11 parameters are needed. They are summarized as

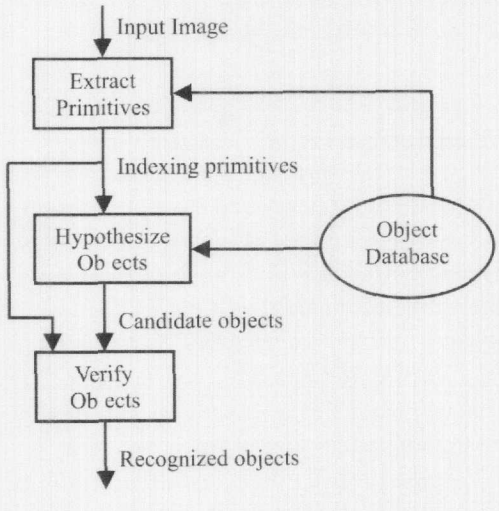
$$\Lambda = (a_1, a_2, a_3, \varepsilon_1, \varepsilon_2, \phi, \theta, \varphi, p_x, p_y, p_z) \tag{2}$$

with $(\phi, \theta, \varphi, p_x, p_y, p_z)$ being the 3D pose parameters.

Parameter recovery methods can be classified into three categories. The first class of methods concentrates on representing single-part objects by assuming that they are pre-segmented (Solina and Bajcsy, 1990). The second category first segments objects of interest into single parts and then represents each part with a superquadric model (Metaxas and Terzopoulos, 1993). The third type directly recovers multiple superquadric models from images without pre-segmentation (Jaklic *et al.*, 2001; Leonardis *et al.*, 1997). We used the second type of methods to first segment a scene into individual objects and then fit superquadrics to each single object.

Figure 6(a) shows a range image of a bin consisting of multiple objects including two blocks, an ellipsoid, a cone, a pyramidical object, and a cylinder. In Figure 6(b), we see a recovered and rendered superquadric for the cone in the bin, and in Figure 6(c) for the pyramid. Both of the recovered superquadrics have accurate parameters when compared with the ground truths of the original objects. At the end of the process, quantitative informations including size, shape, position, and orientation of the object are also obtained. Using this information the system can now compute grasp points and plan the arm path.

Figure 5 Components of a typical object recognition system (Dickinson *et al.*, 1992)



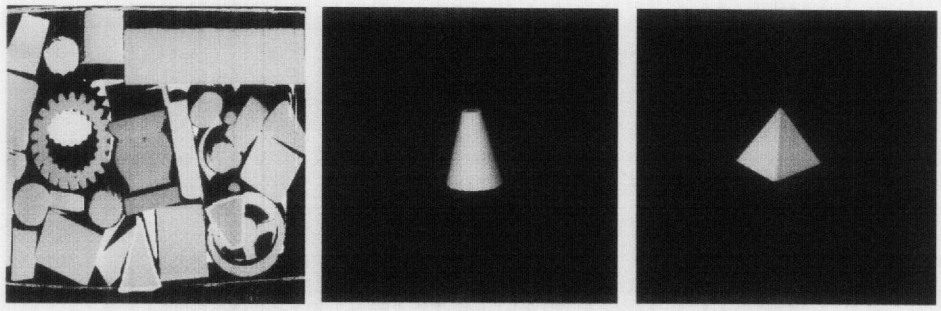
Source: Dickinson *et al*, (1992)

5. Robot arm tracking in video and real-time pose validation

This section describes a method to validate robot arm positions using video sequences. Since the robot arm has to hold the weight of the part, it is possible that the arm will move out of the desired path. This drift can be due to either mechanical reasons or accumulated errors in the sensors. To overcome this problem, we used a high-resolution video camera to monitor the activity of the robot arm and to check its position. This pose validation system is composed of two main modules:

(1) a Gaussian color detector to extract painted color regions on the robot arm, and

Figure 6 Superquadric representation of a bin consisting of multiple objects. (a) Range image of a bin, (b) recovered and rendered superquadric for the cone on the top of the bin, and (c) recovered and rendered 3D superquadric for the pyramidical object in the bin



(a) (b) (c)

(2) a module that estimates the pose for each extracted color region, and compares the regions with a CAD model of the arm.

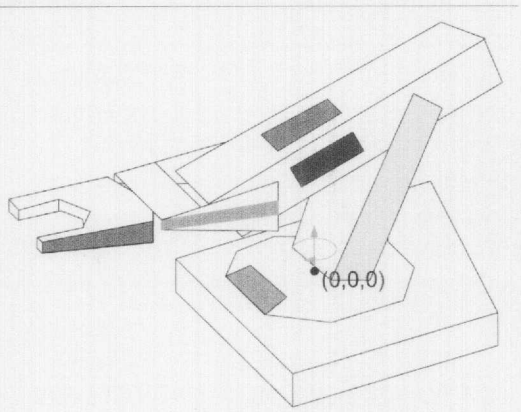
A video camera provides 2D images of 3D world objects, so it is not easy to extract the 3D position information from a single video source in real-time. But if the geometry of the object of interest is known and the camera parameters are available, the position can be recovered using pose estimation methods. Therefore, attaching a known marker, such as rectangular objects of a particular color to the robot arm, will allow real-time tracking. In our investigation we used several rectangles painted in pure colors, which are unlikely to be confused with other scene objects. We also assume that the camera is color-corrected. An example of a set of markers is shown in Figure 7.

The first step for segmenting the painted rectangles in the video images is the building of a multi-rate, Gaussian distribution-based, color detector, which can be expressed in the RGB color space as

$$p(\vec{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \times \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^t\Sigma^{-1}(\vec{x}-\vec{\mu})\right] \quad (3)$$

where $\vec{x}$ is the RGB vector, $\vec{\mu}$ is the mean of RGB values for a given marker, and Σ is a 3×3 covariance matrix. The input value $\vec{x}$ represents the color values for each pixel and the output is the probability of the color. In other words, if $\vec{x}$ is close to the modeled color, then the value of $p(\vec{x})$ will be larger than that of different colors of the spectrum. Using this approach the different markers are reliably extracted from the video stream when assuming fixed *a priori* known illumination with only small variations for indoor environments. In the case of significant changes in illumination, for example when daylight is mixed with artificial light, an additional supervised color correction module has to be included in the system. Once the different markers are extracted, the 3D position of the robot arm is computed relative to the camera's coordinate frame and the grasp path is adjusted accordingly.

Figure 7 Color markers used to determine the pose of the robot arm



6. Conclusion

In this paper, we described the components of a vision system for aiding bin picking tasks. By using both video and range sensors we achieved greater flexibility than in the case of single modality approaches. Range sensors were used for the accurate recovery of parts geometry, while video images served for the tracking and validation of the robot arm position. Superquadric representation of bin objects was applied to grasp point computation and to path planning.

References

Al-Hujazi, A. and Sood, A. (1990), "Range image segmentation with applications to robust bin-picking using vacuum gripper", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 20 No. 6, pp. 1313-24.

Barr, A.H. (1984), "Global and local deformations of solid primitives", *Computer Graphics*, Vol. 18 No. 3, pp. 21-30.

Besl, P.J. and McKay, N.D. (1992), "A method for registration of 3-d shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14 No. 2, pp. 239-56.

Campbell, R. and 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.

Dickinson, S.J., Pentland, A.P. and Rosenfeld, A. (1992), "From volumes to views: an approach to 3-d object recognition", *CVGIP: Image Understanding*, Vol. 55 No. 2, pp. 130-54.

Ikeuchi, K. (1987), "Generating an interpretation tree from a CAD model for 3d-object recognition in bin-picking tasks", *International Journal of Computer Vision*, Vol. 1, pp. 145-65.

Jaklic, A., Leonardis, A. and Solina, F. (2001), *Segmentation and Recovery of Superquadrics*, Kluwer Academic Publishers, Dordrecht.

Kak, A. and Edwards, J. (1995), "Experimental state of the art in 3D object recognition and localization using range data", *Proceedings of the workshop on vision for robots in IROS1995*, Pittsburgh, PA.

Leonardis, A., Jaklic, A. and Solina, F. (1997), "Superquadrics for segmentation and modeling range data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19 No. 11, pp. 1289-95.

Metaxas, D. and Terzopoulos, D. (1993), "Shape and non-rigid motion estimation through physics-based synthesis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15 No. 6, pp. 580-91.

Rahardja, K. and Kosaka, A. (1996), "Vision-based binpicking: recognition and localization of multiple complex objects using simple visual cues", *IEEE Proceedings of International Conference on Intelligent Robots and Systems*, Osaka, Japan, Vol. 3, pp. 1448-57.

Solina, F. and Bajcsy, R. (1990), "Recovery of parametric models from range images: the case for superquadrics with global deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12 No. 2, pp. 131-47.

Tsai, R.Y. (1987), "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", *IEEE Journal of Robotics and Automation*, Vol. 3 No. 4, pp. 323-44.

Further reading

Press, W., Vetterling, W., Teukolsky, S. and Flamery, B. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Press, Cambridge.