Segmentation and Location Computation of Bin Objects

Hema C.R., Paulraj M.P., Nagarajan R. and Sazali Yaacob
Robotics and Intelligent Control Research Group
School of Mechatronics Engineering
Northern Malaysia University College of Engineering,
Kangar, Perlis, Malaysia
Corresponding author E-mail: (hema@kukum.edu.my)

Abstract: In this paper we present a stereo vision based system for segmentation and location computation of partially occluded objects in bin picking environments. Algorithms to segment partially occluded objects and to find the object location [midpoint, x, y and z co-ordinates] with respect to the bin area are proposed. The z co-ordinate is computed using stereo images and neural networks. The proposed algorithms is tested using two neural network architectures namely the Radial Basis Function nets and Simple Feedforward nets. The training results for feedforward nets are found to be more suitable for the current application. The proposed stereo vision system is interfaced with an Adept SCARA Robot to perform bin picking operations. The vision system is found to be effective for partially occluded objects, in the absence of albedo effects. The results are validated through real time bin picking experiments on the Adept Robot.

Keywords: Stereo Vision, Segmentation, Radial Basis Function Nets, Neural Networks, Bin Picking, Object Localization.

1. Introduction

Bin Picking Robot requires information of the object to be picked and its exact location with respect to the bin area. It is generally assumed that the topmost object will be desired object to be picked from a bin with scattered or piled objects. In this paper emphasis is given to solving the segmentation problem of occluded objects in the bin. The stereo vision system proposed considers two aspects, one is the segmentation of the bin image to identify the topmost object and the other is the location of the object midpoint (x, y and z co-ordinates). In bin picking environments object occlusions in the bin image pose a challenge to the segmentation process. Several papers have been published on bin picking algorithms, of which very few have considered occluded objects. Most researches on bin picking use vision only for object recognition and pose determination (Krisnawan Rahardja & Akio Kosaka, 1996), (Ayako Takenouchi & et al,1998), (Ezzet Al-Hujazi & Arun Sood, 1990), (Harry Wechsler & George Lee Zimmerman, 1989), (Kohtaroh Obha & Katsushi Ikeuchi,1996), while others use a model based approach which compares the object image with a model database for pose determination (Yoshikatsu Kimura & et al, 1995), (Martin Berger & et al, 2000), (Sarah Wang & et al, 1994) . Some other approaches use a combination of sensors and model database to solve the bin picking problem (Martin Berger & et al, 2000) who use stereo and CAD models to determine pose of objects.

In contrast to these approaches, a stereo based vision system is proposed which helps the robot to automatically identify and pick the topmost object. The stereo vision system consists of a hardware component and a software component, the hardware comprises of two machine vision cameras in a stereo rig. PULNIX progressive scan cameras are used in this experiment. A stereo base length of 70mm is found to be suitable for the current application; a vacuum gripper is used for grasping the objects. Similar geometrical shaped objects with partial occlusions are used in the experimental process. The software component comprises of segmentation and an object location algorithm.

The proposed segmentation algorithm uses binary thresholding techniques and image histogram to identify the topmost object in the bin as detailed in Section 2. Section 3 elaborates on the object feature extraction process, while Section 4 and 5 describe the neural network architectures and experimental results respectively. Section 6 comprises of the observations, conclusion and future research aspects of the paper.

2. Segmentation

2.1. Image Acquisition and Preprocessing

The stereo images of the occluded bin objects are captured using the stereo cameras, direct lighting of the bin is avoided to reduce brightness and albedo effects. Objects in the bin, partially occlude each other and have
different intensity levels due to their location. The captured images are monochromatic 640 x 480 pixels size. The acquired images are pre-processed to improve the efficiency of the segmentation process. The pre-processing involves two stages i) image resizing and ii) noise filtering. Images are resized to minimize the processing time and to improve the efficiency of the system without significant loss of information of the objects. A image size of 128 x 96 pixels is found to be suitable for the current application. Since the topmost object will be the one without occlusions and the one with higher intensity levels in comparison to the rest of the objects, filtering techniques are applied to smooth out the intensity of the objects and to enhance the edges. The images are filtered using a regional filter and a mask. This masked filter, filters the data in the image with the 2-D linear Gaussian filter and a mask the same size as the original image. This filter returns an image that consists of filtered values for pixels in locations where the mask contains ones and unfiltered values for pixels in locations where the mask contains zeroes. The above process smoothen the intensity of the image around the objects. The resulting filtered image is then subjected to segmenting techniques as detailed in the following section.

2.2 Bin Image Segmentation

Bin image segmentation involves identifying the top most object from the cluster of objects in the bin for pick up. Segmentation of occluded objects has been dealt by Rahardja and Kosaka (Krisnawan Rahardja & Akio Kosaka. 1996), who developed a stereo based algorithm to find simple visual clues of complex objects, this algorithm depends on human assistance. In contrast to this we propose an automatic segmentation algorithm which can segment partially occluded bin objects. Since all the objects are partially occluded except the topmost object, separating the topmost object can be done using the grey value of the object. A histogram of the bin stereo images displays the grey levels of the image. Segmentation using binary thresholding is possible by identifying the pixels of grey levels higher than a threshold value which is assumed to relate to the topmost object, as the topmost object has a brightness level higher than the other objects in the bin. A suitable threshold segmenting only the topmost object is to be computed.

For real-time bin picking, automatic determination of threshold value is an essential criterion. To determine this threshold value an algorithm is proposed which uses the grey levels of the image from the histogram of both the stereo images to compute the threshold. The proposed algorithm is as follows:

**Segmentation Algorithm**:

Step 1: The histogram is computed from the left and right gray scale images for a bin value of 0 to 255. Counts a (i), i=1,2,3,...,256 contains the number of pixels with a gray scale value of (i-1) pixels for the left image.

Counts b (i), i=1,2,3,...,256 contains the number of pixels with a gray scale of (i-1) for the right image.

Step 2: Compute the logarithmic weighted gray scale value of the left and right image as

\[ t_a (i) = \log(\text{count } a (i)) \times (i-1) \]  
\[ t_b (i) = \log(\text{count } b (i)) \times (i-1) \]

where \( i = 1,2,3,...,255 \)

Step 3: Compute the logarithmic weighted gray scale

\[ t_m = \frac{1}{256} \sum_{i=1}^{256} t_a(i) \]  
\[ t_mb = \frac{1}{256} \sum_{i=1}^{256} t_b(i) \]

Step 4: The threshold T is the minimum value of ‘tm’ and ‘tmb’.

Threshold of both the stereo images are computed separately and the min value of the two thresholds is applied as the threshold to both the images.

The topmost object is segmented and converted into a binary image. The performance of the proposed algorithm was compared with the more popular global thresholding algorithm proposed by Otsu. Fig.1 shows the results of the comparison of both the methods.

![Fig. 1 Comparison of thresholding methods](image-url)
3. Object Location

The next phase is to compute the location of the object with respect to the bin area. To pick up the topmost object the robot must be provided with the location co-ordinates of the object midpoint [x, y and z co-ordinates]. The x and y co-ordinates are determined from the 2D segmented image of the object, whereas the z co-ordinate which is the depth or distance co-ordinate requires the 3D representation. To find the z co-ordinate a neural network and stereo image approach is proposed which trains a neural network to compute the distance or z co-ordinate of the object from its stereo images. Two neural nets namely radial basis function nets and feedforward nets are compared in computing the z co-ordinate. The object features of the stereo images are used as the input data and the distance of the object from gripper is used as the output data to train the neural network. The object features are extracted using singular value decomposition [SVD]. The following sections elaborate on SVD and feature extraction process.

3.1 Singular Value Decomposition

The Singular Value Decomposition (SVD) is a widely used technique to decompose a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix (Zi-Quan Hong, 1991). Any \( m \times n \) matrix \( A \) \((m \gg n)\) can be written as the product of a \( m \times m \) column-orthogonal matrix \( U \), an \( m \times n \) diagonal matrix \( W \) with positive or zero elements, and the transpose of an \( n \times n \) orthogonal matrix \( V \). The diagonal elements of matrix \( W \) are the singular values of matrix \( A \), which are non-negative numbers. The singular values obtained by the SVD of an image matrix give algebraic features of an image, which represents the intrinsic attributes of the image (Zi-Quan Hong, 1991).

3.2 Feature Extraction

To pick up the topmost object, the location of the object is to be computed. Since only one object is present in the segmented image, finding the centroid of the object will provide the x and y coordinates of the object midpoint, however it is also important to know the z coordinate [distance] to move the gripper for pick up process. Stereo feature matching is one of the popular methods for distance computation. In this research we present a method to compute the distance from the added stereo images of the object, called unified stereo imaging, wherein the stereo images of the object are added for feature extraction. The features of the added image varies with respect to the distance of the object, this aspect is used to train the neural network. To optimise the computational time edge images of the added image is used in the feature extraction process. Singular value features are extracted using SVD, for training the network singular values are extracted for various distance of the object. Fig. 2 shows the flow diagram for object feature extraction. Singular values feature having values below ‘1’ are ignored as few prominent singular value features of the object are sufficient to represent the object distance (Zi-Quan Hong, 1991). Hence ten significant features of the object image are fed to a neural network to train the network for object distance computation.

![Fig. 2. Flow Diagram for Object Feature Extraction](image-url)
3.3 Radial Basis Function Network Architecture

An extremely powerful type of feedforward artificial neural network for pattern recognition is the Radial Basis Function [RBF] Network, which differs strongly from the multiple layer networks in the activation functions. The architecture consists of three layers: an input layer, a hidden layer of \( j \) basis functions and an output layer of \( K \) linear output units. The activation values of the hidden units are calculated as the closeness of the input vector \( X_p \) to an vector \( \mu_j \) associated with hidden unit \( Z_j \). The activation is given as

\[
Z_{j,p} = e^{-\frac{\| X_p - \mu_j \|^2}{2\sigma_j^2}} \tag{5}
\]

where \( \sigma_j \) and \( \mu_j \) are respectively the standard deviation and mean of the basis function. The final output is calculated as

\[
Y_{k,p} = \sum_{j=1}^{J} w_{kj} Z_{j,p} \tag{6}
\]

Fig. 3 illustrates a RBF Network architecture. The \( \mu_j \) vector is the weight vector to the \( j \)-th hidden unit. (Basis function) and \( \sigma_j^2 \) is the bias to the \( j \)-th hidden unit also known as the spread factor. In the case of classification problems an RBF network finds the centroid of data clusters and uses this centroid as the centers of the Gaussian density function.

3.4 Simple Feedforward Network Architecture

The simple feedforward neural network architecture consists of three layers. 10 singular values are fed to the network as input data. The hidden layer is chosen to have 5 neurons and the output consists of 1 neuron, which represents the object distance. The hidden and input neurons have a bias value of 1.0 and are activated by bipolar sigmoid activation function. The initial weights for the network are randomized between -0.5 and 0.5 and normalized. The initial weights that are connected to any of the hidden or output neuron are normalized in such a way that the sum of the squared weight values connected to a neuron is one. This normalization is carried out using equation (7), which is used to implement the weight updating (S.N.Sivanandam & M.Paulraj, 2003).

\[
w_{ij}^{(new)} = \frac{w_{ij}^{(old)}}{\sqrt{w_{i1}^2 + w_{i2}^2 + \ldots + w_{ip}^2}}, j=1,2,3,\ldots,p \tag{7}
\]

where

- \( n \) - number of input units
- \( p \) - number of hidden units

A sum squared error criteria as defined by equation (8) is used as a stopping criteria while training the network. The sum-squared tolerance defined in equation (8) is fixed as 0.01. The network is trained by the conventional back propagation procedure (Laurene Faussett, 1990). The cumulative error is the sum squared error for each epoch and is given by:-

\[
\text{Sum squared error} = \sum_{p=1}^{n} \sum_{k=1}^{m} (t_k - y_k)^2 \tag{8}
\]

where

- \( t_k \) is the expected output value for the \( k \)-th neuron,
- \( y_k \) is the actual output value for the \( k \)-th neuron
- \( m \) is the total number of output neurons, and
- \( p \) is the total number of input neurons.
4. Experimental Phase

In the experimental phase, 53 images were acquired. The bin images are pre-processed to smooth the intensity level of the object. The pre-processed images are segmented to extract the topmost object in the bin. The segmented left and right images are added and the edge of the added image is extracted. The ‘x’ and ‘y’ co-ordinates are computed from the left segmented object image, which is considered as the reference image, by finding the centroid of the reference image. To compute the ‘z’ co-ordinate, the singular value features are extracted from the added edge image using SVD. Ten singular values of an image are fed as input and the object distance is fed as output to the neural networks. The feature data is used to compare the performance of the two networks.

The RBF network is trained with 40 sample data, the spread factor \( \sigma_j \) is chosen as 3.2. The network trained in 1.407 seconds. Subsequently the network is tested for 53 samples. The radial basis function network successfully computed the distance for 47 samples out of 53 samples, resulting in computation efficiency of 88.6%. The training parameters and results are tabulated in Table 1.

| Neural Network Parameters | Values |
|---------------------------|--------|
| No. of input neurons      | 10     |
| No. of Output neurons     | 1      |
| Learning Rate             | 0.1    |
| Tolerance                 | 0.001  |
| Training time             | 70.7 seconds |
| Training Samples          | 40     |
| Test Samples              | 53     |
| Success Rate              | 92.45% |

Table 1: Training Parameters of the RBF Neural Network

The simple feedforward network is trained with 40 images samples using the backpropagation training procedure. In the testing phase the network is tested with 53 sample data. The proposed method is found to successfully compute the distance for 49 images with a success rate of 92.45% and error tolerance of 0.001. Table 2 shows the training parameters and test results of the neural network.

| Neural Network Parameters | Values |
|---------------------------|--------|
| No. of input neurons      | 10     |
| No. of Output neurons     | 1      |
| Spread Factor             | 3.2    |
| Training time             | 1.407 seconds |
| Training Samples          | 40     |
| Test Samples              | 53     |
| Success Rate              | 88.6%  |

Table 2: Training Parameters of the Simple Feedforward Network

In the real time experimentation phase, the developed stereo vision system is interfaced with an Adept SCARA 600 Robot. A vacuum gripper is used for pick and place operation. A bin with seven partially occluded objects is used for testing. The Adept SCARA Robot is tested for real time bin picking using the object location co-ordinates \([x', y' \text{ and } z']\) computed by the stereo vision system. The bin picking system is designed to systematically pick the topmost object one at a time. The process is repeated till all the objects in the bin are picked. At an average, six out of the seven objects were picked and placed successfully with a bin pick and place performance of 85.7%.

5. Conclusion

A bin picking system using stereo vision sensors for object localization is presented. Algorithms for segmentation of partially occluded bin objects and location of topmost object is proposed. The results of two networks, RBF and feedforward nets are compared in evaluating the Z co-ordinate of object midpoint. The feedforward net is found to be more suitable for the current application. The system is experimentally verified and real time bin picking results are presented. Real time tests validate the applicability of the proposed algorithms for bin picking. The major constraint of the proposed system was poor segmentation in the presence of albedo effects and uneven brightness in certain parts of the bin. Optimal lighting conditions are essential to derive satisfactory results. Future work will include optimising lighting conditions and improving the network performance.

6. References

Ayako Takenouchi, Naoyoshi Knamaru and Makoto Mizukawa. (1998). Hough-space-based Object Recognition tightly coupled with Path planning for Robust and Fast Bin-picking, Proc. IEEE /RSJ International Conference on Intelligent Robots and Systems, Canada, pp:1222-1228.

Ezzet Al-Hujazi & Arun Sood. (1990). Range Image Segmentation with Applications to Robot Bin-Picking Using Vacuum Gripper, IEEE Transactions on System, Man and Cybernetics, vol. 20. No.6, pp:1313–1325

Harry Wechsler & George Lee Zimmerman. (1989). Distributed Associative Memory (DAM) for Bin-Picking, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. II. No. 8, pp:814-822.

Kohtaro Ohba & Katsushi Ikeuchi. (1996). Recognition of the Multi Specularity Objects for Bin Picking Task, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, vol: 3, pp:1440 – 1447.

Krisnawan Rahardja & Akio Kosaka. (1996). Vision-Based Bin –Picking: Recognition and Localization of Multiple Complex Objects Using Simple Visual Cues,
Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, October, vol: 3, pp. 1448 – 1457.
Laurene Fausett. (1990). Fundamentals of Neural Network Architecture and Applications, Prentice Hall, USA.
Martin Berger, Gernot Bachler and Stefan Scherer. (2000) Vision Guided Bin Picking and Mounting in a Flexible Assembly Cel, Thirteenth International Conference on Industrial and Engineering Application of Artificial Intelligence and Expert Systems. U.S.A.
Sarah Wang, Robert Cromwell, Avi kak, Ichiro Kimura and Michiharu Osada, (1994) Model-Based Vision for Robotic Manipulation of Twisted tubular parts: Using affine transforms and heuristic Search, IEEE Transactions, pp:208 – 215.
S.N.Sivanandam & M.Paulraj. (2003). Introduction to Artificial Neural Networks Vikas Publishing House, India.
Yoshikatsu Kimura, Takashi Naito, Masaru Nakano. (1995). Stereo Vision System for Car Assembly. IEEE International Conference on Robotics and Automation, pp. 1403-1409.
Zi-Quan Hong. (1991). Algebraic Feature Extraction of Image for Recognition, IEEE Transactions on Pattern Recognition, vol. 24 No. 3 pp: 211-219.