Segmentation and Object Recognition for Robot Bin Picking Systems

R. Nagarajan[+], Sazali Yaacob*, Paulraj Pandiyan[+], Hema C.R.[+], Shamsudin Amin[++], Marzuki Khalid[++], M. Rizon*

[+]*Artificial Intelligence Group*
School of Engineering and Information Technology
Universiti Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia
[++]*Center for Artificial Intelligence and Robotics*
Faculty of Electrical Engineering
Universiti Teknologi Malaysia Kuala Lumpur, Malaysia
******School of Mechatronic Engineering*
Northern Malaysia University College of Engineering
01000 Kangar, Perlis, Malaysia
E-mail: rizon@kukum.edu.my

Abstract

This paper presents a vision based segmentation and object recognition system for bin picking applications. Objects in the bin are usually placed in a disorganized manner inside the bin. A method to segment the topmost object from the bin image is presented using threshold derived from the histogram of the bin image. Once the object is segmented its singular value features are extracted for object recognition. Object recognition is accomplished using a simple feed forward neural network. The proposed method is capable of segmenting occluded object as well as recognizing complex objects.

Keywords: Bin Picking, Object Recognition, Occluded Object Segmentation, Singular Value Decomposition.

1 Introduction

In bin picking robot systems, objects to be picked are placed inside a bin in an unorganized pile. Any vision system attached to a bin picking robot must be able to analyze the image of the bin and be able to locate all the parts seen as well as identify them correctly. Research on bin picking systems had started in the 80s' but most bin picking publications consider only non-occluded objects. Some researches pick up an object and analyse the object for pose determination and recognition using database of the object images, while others rely on sensors other than the vision sensors[1],[2],[3]. We present here a vision based system to isolate and recognise the object to be picked. If all the parts are separated from each other, then identifying the parts is relatively easy. However when the parts are partially occluded more than one object will be merged into a single blob making it difficult to identify each one of them.

The image segmentation that divides the image into meaningful sub regions is an indispensable step in image analysis. One approach of solving the problem of identifying objects is by detecting and measuring all segments belonging to the object border and all internal angles [4]. A regional growth is a technique that starts at the known pixel points and extends to all neighbouring pixels that are similar in gray level, colour texture, or other properties in order to form a complete region. Using the difference chain code in contour encoding is another way of recognizing partially occluded object. An object shape is identified by the system through the detection of selected discrete feature segments in the contour code instead of attempting to search for a complete boundary [5]. Image segmentation can be considered as a clustering process in which the pixels are classified to specific regions based on their gray-level values and spatial connectivity.

We propose a vision based segmentation and object recognition method for occluded objects using thresholding, singular value decomposition and Neural Networks. Section 2 briefly introduces singular value decomposition. Section 3 describes the methodology of segmentation that includes feature extraction and a neural network architecture for object recognition. The experimental results and conclusion derived from this work are given in section 3 and 4 respectively.

2 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 [6]. Any 'm x n' matrix **a** ($m >= n$) can be written as the product of a 'm x m' column-orthogonal matrix **u**, an 'm x n' diagonal matrix **w** with positive or zero elements, and the transpose of an 'n x n' orthogonal matrix **v** [7]:

$$a = u\, w\, v^{t} \qquad \ldots (1)$$

where

$$\mathbf{w} = \begin{bmatrix} w_1 & 0 & . & 0 & 0 \\ & w_1 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & w_{n-1} & 0 \\ 0 & 0 & . & 0 & w_n \end{bmatrix} \qquad (2)$$

and

$$u^t u = vv^t = I \qquad (3)$$

where

$w_1, w_2, ..., w_{n-1} w_n \geq 0$,

and 't' is the transpose.

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 is an algebraic feature of an image which represents the intrinsic attributes of an image[6].

3 Methodology

Images of the occluded objects in the bin are captured using a Pulnix TM 6702 CCD machine vision camera. Direct bin lighting of the bin is avoided to reduce brightness and albedo effects. The captured images are of size 640 x 480 pixels and are monochromatic. To minimise the processing time and to improve the efficiency of the system without significant loss of information of the objects, the images are resized to 128 x 96 pixels. The resized images have to be further processed using filtering techniques to identify the borders of the objects and to smooth the intensity of the object region. First the image is 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 for filtering. This filter returns an image that consists of filtered values for pixels in locations where the mask contains 1's, and unfiltered values for pixels in locations where the mask contains 0's. The above process smoothens the intensity of the image around the objects. The resultant image is passed through a median filter which defines the object edges. The median filter performs median filtering of the image matrix in two dimensions. Each output pixel contains the median value in the M-by-N neighbourhood around the corresponding pixel in the input image. The filter pads the image with zeros on the edges, so that the median values for the points within [M N]/2 of the edges may appear distorted. The M-by- N is chosen according to the dimensions of the object. In our experiment, a 4 x 4 matrix was chosen. The resulting filtered image is then subjected to segmenting techniques as detailed in the following section.

3.1 Bin Image Segmentation

Bin image segmentation involves identifying the top most object from the cluster of objects in the bin for picking up. 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 images displays the grey levels of the image. Pixels of grey levels higher than a threshold value relate to the topmost object. The threshold value is determined by taking log of the magnitudes using equation (4) for 40 maximum values.

$$T = [\log10\ (\text{COUNTS a})]\ i \qquad (4)$$

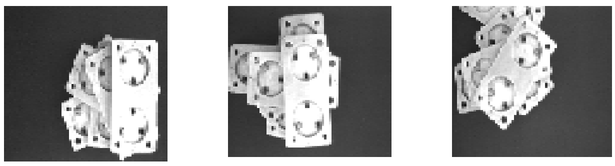
where

'T' is the threshold

'i' is the Gray scale value

'COUNTS a' is the Gray value count

The maximum value is chosen experimentally to get satisfactory results for all images. The grey images are converted to binary images by applying the threshold derived from the histogram using equation (4). The Figure 1 shows the bin images before and after segmentation.



(a) Bin images

(b) Segmented images showing the top most object

Figure 1 Bin Image Segmentation

3.2 Object Feature Extraction

The topmost object extracted from the segmentation process varies in orientation and pose as the objects in the bin are not placed in an organised pile. It now becomes essential to recognise the object irrespective of its orientation. To recognise the object, features of the object in various poses and orientation are extracted to train a neural network to recognise the object. In the feature extraction process the edge images of the segmented object is extracted using a canny edge detector. The edge image matrix is decomposed using singular value decomposition to extract the singular value. Edge features minimise the computational time and pocess optimum information of the object for recognition. Figure 2 shows the edge images of object with different orientation and pose.

Figure 2 Edge images of object

The singular value features extracted from the edge images are the object features; singular values below 1 are ignored since they h are sometimes due to noise in the image and do not relate to the object under consideration. Twenty five such significant features of the object is fed to a neural network to train the network for object recognition.

3.3 Neural Network Architecture

The Neural Network architecture shown in Figure 3 consists of three layers. 25 singular values are fed to the network as input data. The hidden layer is chosen to have 13 neurons and the output consists of 2 neurons which represent the two objects to be recognised. The hidden and input neurons have a bias value of 1.0 and are activated by binary sigmoid activation function. The choice of initial weight will influence whether the network reaches a global minimum of the error and if so how quickly it converges. It is important that the initial weights must not be large in order to avoid the initial input signals to each hidden or output unit falling in the saturation region. On the other hand, if the initial weights are too small, the net input to a hidden or output unit will be close to zero which also causes extremely slow learning [8]. Hence the initial weights for the above 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 (5) which is used to implement the weight updating.

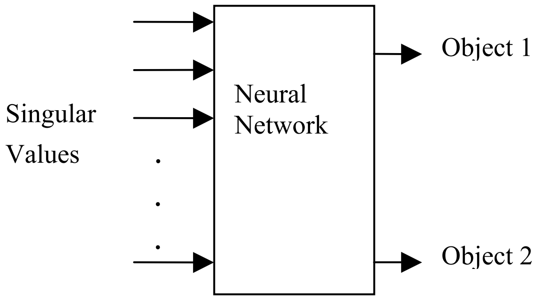


Figure 3 A Feed Forward Neural Networks for Object Recognition

$$w_{1j}(new) = \frac{w_{1j}(old)}{\sqrt{w_{1j}^2 + w_{2j}^2 + \ldots + w_{nj}^2}} \quad , j=1,2,3,\ldots,p \qquad (5)$$

where n - number of input units

p - number of hidden units

A sum squared error criteria as defined by equation (6) is used as a stopping criteria while training the network. The sum squared tolerance defined in equation (6) is fixed as 0.01. The network is trained by the conventional back propagation procedure. The cumulative error versus epoch plot of the trained neural network is shown in Figure 4. The cumulative error is the sum squared error for each epoch and is given by:-

$$\text{Sum squared error} = \sum_{p=1}^{p} \sum_{k=1}^{m} (t_k - y_k)^2 \qquad (6)$$

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 Results

In the experimental study two sets of bin images are acquired for two different objects. Only similar objects are placed in the bin at a time. The neural network is trained with 60 percent of the data and tested with 116 image data. Table I lists the training parameters of the Neural Network. The proposed method is found to successfully classify 93 % of the tested images. Figure 4 shows the Cumulative error versus epoch plot of the neural network. At each iteration, an objective function is minimized to find the best location for the clusters.

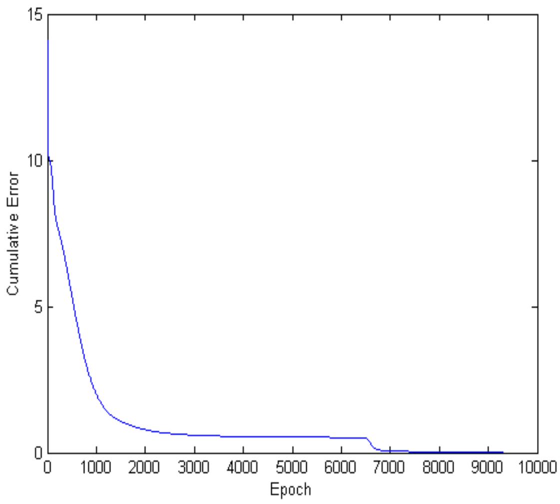


Figure 4 Cumulative error versus Epoch Plot

Table I – Training Parameters of NN

Parameters	Values
No. of input neurons	25
No. of Output neurons	2
No. of hidden neurons	13
Bias value	1.0
Tolerance	0.01
Percentage Classification	93%
Maximum Epoch	9331

4 Conclusion

A vision based segmentation and object recognition method is presented. The method is experimentally verified and results presented are satisfactory. The proposed method however was found to be ineffective for object images with albedo effects and over brightness. Optimal lighting is essential to derive satisfactory results. The proposed method has successfully segmented partially occluded objects. Methods to improve the segmentation results are under study.

Acknowledgement

The authors thankfully acknowledge the funding [Code. No. 04-02-10-0024-PR001] received from IRPA through UTM and UMS.

References

[1]Ayako Takenouchi, Naoyoshi Kanamaru and Makoto Mizukawa, Hough-space-based Object Recognition Tightly Coupled with path planning for Robust and fast Bin-picking, *Proceedings of International Conference on Intelligent Robots and Systems*, Canada, October 1998, pp 1222 – 1229.

[2]Kohtaro Ohba, Katsushi Ikeuchi, Recognition of the Multi Specularity Objects for Bin-picking Task, *Proceedings IROS*, IEEE October 1996 pp 1440 – 1447.

[3]Yoshimasa YANAGIHARA, Toshiro KITA, Parts Picking in Disordered Environment, IEEE International Workshop on Intelligent Robots and Systems 1991, Osaka , Japan pp 517-522.

[4] Paulo Lopes, Eugenio Oliveira. Identification of Partially Occluded Objects Using Object Chain Code. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems Yokohama*, Japan 1829-1834 July 1993

[5]Peter W.M. Tsang, P.C.Yuen. Recognition of Partially Occluded Objects. *IEEE Transactions on Systems, Man and Cybernetic*s Vol. 23, No.1 January February 1993

[6]Zi-Quan Hong, Algebraic Feature Extraction of Image for Recognition *IEEE Transactions on Pattern Recognition* Vol. 24 No. 3 pp 211-219, 1991.

[7] David S. Watkins, *Fundamentals of Matrix Computations*. John Wiley & Sons 2001.

[8]S.N. Sivanandam, M.Paulraj, *Introduction to Artificial Neural Networks* Vikas Publishing House, India. 2003

[9] Laurene Fausett, *Fundamentals of Neural Network Architecture and Applications*, Prentice Hall, USA. 1990