BIN OBJECT RECOGNITION USING IMAGE MATRIX DECOMPOSITION AND NEURAL NETWORKS

Hema C R, Paulraj M., R. Nagarajan, Sazali Yaacob

Robotics and Intelligent Control Research Group
School of Mechatronics Engineering
Northern Malaysia University College of Engineering
Kangar, Perlis, Malaysia
Email: hema@kukum.edu.my

Abstract

Bin picking robots require vision sensors capable of recognizing objects in the bin irrespective of the orientation and pose of the objects inside the bin. Bin picking systems are still a challenge to the robot vision research community due to the complexity of segmenting of occluded industrial objects as well as recognizing the segmented objects which have irregular shapes. The problem becomes more complex when these objects look like entirely different objects in various orientations. In this paper a simple object recognition method is presented using singular value decomposition of the object image matrix and a functional link neural network for a bin picking vision system. The results of the functional link net are compared with that of a simple feed forward net. The network is trained using the error back propagation procedure. The proposed method is robust for recognition of objects.

Key Words: Object Recognition, Neural Networks, Bin Picking, Singular Values

تمييز الأشياء داخل صندوق باستخدام تحليل منضومة ألصوره والشبكات ألعصبيه

سي آر هيما و أم باولراج و آر ناكاراجان و سازالي يعقوب
ألمجموعه ألبحثيه للروبوتيات والتحكم الذكي
مدرسة هندسة الميكاترونكس، الكلية الهندسية الجامعه لشمال ماليزيا،
كانكار، بيرلايس، ماليزيا

تتطلب منضو مات الالتقاط من الصناديق استعمال منضو مات رؤية قادرة على تمييز الأشياء داخل الصندوق. تصعب عملية التمييز هذه وذلك لأن الأشياء المراد تمييزها تأخذ أشكالا غير منتضمه. كما إن تغير وضعية اللاجسام داخل الصندوق يؤدي الى رؤيتها كأجسام مختلفة. يقدم هذا البحث إلى طريقه مبسطه لتمييز الأجسام داخل الصناديق بطريقة تحليل مصفوفة الصورة وشبكه عصبيه ذات داله مرتبطة. تمت مقارنة النتائج مع تلك المستحصله من شبكه عصبيه بسيطة ذات تغذيه أمامية. تم تعليم ألشبكه باستخدام خوارزمية انسياب الخطأ خلفا. وقد أضهرت النتائج أن المنضومه المفترحه ذات متانة عالية.

1.0 INTRODUCTION

Mechanical manipulators are being used increasingly for machine loading, welding, painting and sealing. However they have not been used extensively for applications such as assembly operations. One of the reasons for this is that the manipulators typically just play back a sequence of motions. The blind robot has difficulty in dealing with uncertainty in the position of objects [1]. Feeding mechanism and fixtures are needed to present the objects precisely in predefined locations for pick up, thereby increasing the cost of automation of such operations. The above

problem can be tackled with a bin picking system wherein the objects to be picked are placed inside the bin in an unorganized manner. The bin picking system with vision sensors depicts a manual assembly unit where objects to be assembled are placed in bins surrounding the workstation. However the vision system attached to the bin picking system
should be capable of analyzing the objects in the bin as well as identify the object to be picked up from the pile irrespective of the pose or orientation of the object. In this paper we address the problem of recognizing complex industrial objects with different orientations.

Many researchers have approached bin object recognition in different ways. A model-based and stereo approach. is presented by Martin Berger and et.el [2] wherein the pose of the object is compared with a CAD model of the object for recognition. Another approach is to track and fixate the grasping position of an object is presented by Nan-Feng and et-el [3]. Recognition of partially occluded objects for bin picking tasks using Eigen space analysis is reported by Obha and Ikeuchi [4] the method uses global similarity among eigen windows to select the windows beyond a threshold this eliminates mismatch of wrong windows. Image windows are compared with dictionary images to recognize the objects. Most of the research on bin picking concentrates on ranges finders, structured light and CAD models to recognize the objects. We present a method for object recognition, which is very different from the methods seen above. The method proposed uses the images of industrial objects in various orientations, acquired by a digital camera to extract the singular value features. These features are fed as input to a functional link neural network classifier [FLNN], which identifies the objects.

The results of the proposed method using FLNN classifier is found to be better in terms of classification and training time in comparison with a conventional feed forward neural classifier [5].

2.0 FEATURE EXTRACTION

Feature extraction is the most fundamental and important problem in the field of object recognition [6] and finding the efficient feature is always the key to solving the problem of object recognition. There are four major types of features for object recognition.

(a) *Visual Features*: They include edges, contours, textures and regions of an image; these are all visual features of a pixel.

(b) *Statistical Features*: Histogram features, properties of moments, belong to this kind of features of a pixel.

(c) *Transform Coefficient Features*: Fourier descriptors have a good ability to describe edges of objects and Fourier transforms are used to extract texture features of images.

(d) *Algebraic Features*: They represent intrinsic attributions of an image. Any image can be considered as a matrix; therefore various algebraic transforms or matrix decompositions can be used for algebraic feature extraction of the image.

Eigenvector of a matrix represents algebraic attributes of the matrix and hence can be used for feature extraction. Similarly Singular Value Decomposition [SVD] of a matrix results in singular values, which is another algebraic feature of an image. The 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 [7]. Any m x n matrix $\boldsymbol{A}$ ($m >= n$) can be written as the product of an m x m column-orthogonal matrix $\boldsymbol{U}$, an m x n diagonal matrix W with positive or zero elements, and the transpose of an n x n orthogonal matrix $\boldsymbol{V}$ [8]:

$$A = U\,WV^{t} \tag{1}$$

The singular values [SV] obtained by the Singular Value Decomposition of an image matrix is an algebraic feature of an image which have the properties such as [6]

(a) *Stability:* The SV features do not have large changes when the grey values of an image have small changes.

(b) *Invariance:* The SV features are invariant to proportional variance of image intensity in optimal discriminate space. This property is important for description and recognition of objects.

(c) The invariance of SV features to transposition transforms.

(d) The invariance of SV features to rotation transforms.

(e) The invariance of SV features to translation

(f) The invariance of SV features to mirror transforms

The above properties of the SV features play an important role in the recognition of objects in different orientation.

3.0 FUNCTIONAL LINK NEURAL NETWORKS

Functional link enhances the computing power of the neural network and by functional transformation an input pattern is enhanced in its representation. The neural network saves time to learn additional information obtained via such transformation. The functional link can improve the network in both learning capacity and efficiency.

The functional link generates a set of linearly independent functions and evaluates these functions with the input pattern as the argument. Suppose each input node encodes a certain feature, applying the functional link to a feature causes it to multiply and generate more features. As a consequence the expressive power of the network increases, as does its modeling capability. The benefit of the functional link when applied to mathematical modeling is the increased accuracy of mapping through expansion of the basis set. [9].

The FLNN architecture consists of three layers; the input layer [including the functional link], hidden layer and the output layer. The network architecture is shown in Fig 1.

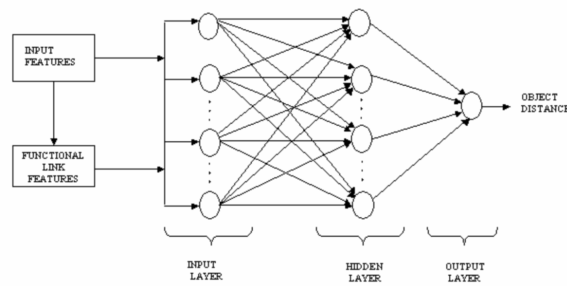

Fig. 1. Functional Link Neural Network

4.0 METHODOLOGY

Images of industrial objects are acquired in different orientation using a Sony DSC P72 digital camera. The acquired images are of size 1024 x 1024 pixels and are preprocessed for image enhancement, binarization and edge detection for the extraction of singular value features. The extracted features are fed as input to the FLNN classifier.

4.1 Image Preprocessing

Images acquired are of 1024 x 1024 pixels size; this size is not convenient for preprocessing, in addition due to lighting contrast the images are accompanied by noise. Hence the image has to be resized to a convenient scale and preprocessed for noise removal. Image enhancement is performed to process the image for noise removal. The object images are resized to 32 x 32 pixels. The image is then passed through a median filter. A median filter is used because these filters are effective in the presence of impulse noise. The image is sharpened to define the edge details using an ideal high pass filter.

4.2 Binarization

A suitable gray level threshold is applied to the sharpened image to compute the global threshold level to convert the intensity image to a binary image. The binary conversion is done to convert the RGB image to grayscale image. The output binary image has values of 0 for all pixels less than the threshold level and 1 for all pixels greater than the threshold level. A clear definition of the image is obtained from which the edge boundary of the object can be extracted.

4.3 Edge Detection

Edge boundary of the image is detected by means of a canny edge detector .This method finds edges by looking for local maxima of the gradient of the image. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be "fooled" by noise, and more likely to detect true weak edges. The singular values can be extracted by the decomposition of the edge image matrix using SVD. SVD produces a diagonal matrix, of the same dimension as the edge image and with nonnegative diagonal elements in decreasing order. The diagonal matrix elements otherwise called as singular values represent intrinsic attributes of the image

5.0 NEURAL NETWORK ARCHITECTURE

The Functional Link Neural Network architecture consists of a feed forward neural network with the functional link inputs. 25 singular values are fed to the network as input data in addition to 49 functional link inputs. The hidden layer is chosen to have 12 neurons and the output consists of 4 neurons. The hidden and input neurons have a bias value of 1.0 and are activated by bipolar sigmoidal activation function. The choice of initial weight will influence whether the net 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.

Otherwise the initial input signals to each hidden or output unit will be likely to fall in the saturation region. On the other hand if the initial weights are too small, the net input to a hidden layer or output layer will be close to zero which can cause extremely slow learning [10]. 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 1. The sum squared error criteria is used as a stopping criteria while training the network [11]. The sum squared tolerance is fixed as 0.01. The network is trained by the conventional back propagation procedure. The cumulative error is the sum squared error for each epoch.

6.0 EXPERIMENTAL RESULTS

In the experimental study, four industrial assembly objects are taken. The images of the objects are captured in different orientations. Each object is captured in five to seven such stable poses [orientation] depending on the shape of the object. Thirty one images of these four objects are acquired. The images are pre-processed as detailed in section 4.1 and 4.2. From the edge image matrix the singular value features are extracted through singular value decomposition and the data are normalized. Out of the thirty two singular values obtained, only first twenty five prominent singular value features of magnitude greater than one are considered for classifying each image. The network is trained for 25 feature data and tested with 31 samples. Fig.2 shows the four objects and samples of some of their orientations.

The orientation of the object is such that all prominent edge features are visible to get a complete picture of the shape of the object. This is essential to get a good description of the shape of a particular object.

The network is trained for 27 sample images and tested for 31 samples. The FLNN is trained using the error back propagation procedure. The success rate of the object classification using the FLNN is 90.32 %. These results were compared with a simple feed forward three layer network classifier and it was observed that the FLNN is able to classify the samples in lesser time and minimum number of epochs as against the time of 53 seconds and 9331 epochs of the former net for the same input parameters and training procedure. A comparison of the training parameters and test results of the proposed FLNN classifier a Conventional NN classifier are shown in Table 1.

7.0 CONCLUSION

In this paper, an object recognition methodology for bin picking robots based on shape classification using the singular value decomposition of the image matrices and FLNN classifier is presented. The successful classification percentage is 90.32% with a tolerance of 0.01.

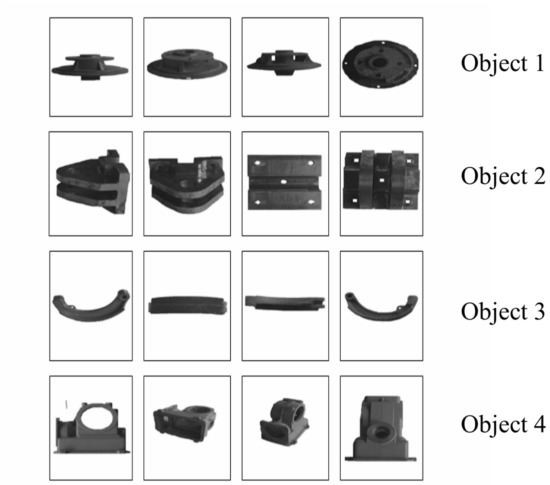

Fig. 2. Sample images of the industrial objects with different orientation

The method proposed using functional link network was able to classify the objects with lesser training time and epoch in comparison with a simple feed forward net classifier. Experiments prove that a functional link neural classifier is more efficient than conventional neural classifiers. Future research will involve improving the classification rate of the classifier.

Table 1. Comparison of Results of FLNN and Conventional Classifier

Network Parameters	FLNN	Conventional NN
No. of input neurons	25	25
No. of Output neurons	4	4
No. of Functional link input neurons	49	nil
No. of hidden neurons	12	12
Bias value	1.0	1.0
Learning Rate	0.1	0.1
Tolerance	0.01	0.01
Training time	6 seconds	53 seconds
Percentage Classification	90.32%	89.28%
Maximum Epoch	5892	9331

REFERENCES

[1] Berthold Klaus Paul Horn, *Robot Vision,* the MIT Press, McGraw-Hill Book Company, London, 1993

[2] Martin Berger, Gernot Bachler and Stefan Scherer.`Vision Guided Bin Picking and Mounting in a Flexible Assembly Cell'. *Thirteenth International Conference on Industrial and Engineering Application of Artificial Intelligence and Expert Systems.* U.S.A. 2000.

[3] Nan-Feng XIAO, Saeid Nahavandi. An Active Stereo Vision Based learning Approach for Robotic Tracking. Fixating and Grasping Control*, IEEE, ICIT '02.* Thailand, 2002.

[4] Kohtaro Ohba and Katsushi Ikeuchi, Recognition of the Multi Specularity Objects for Bin Picking Task, IEEE *IROS'96 proceedings* pp 1440 – 1447,1996

[5] R Nagarajan, Sazali Yaacob, Paulraj Pandiyan, Shamsudin Amin, Marzuki Khalid and Hema C R, Feature Extraction using Singular Value Decomposition for Object Recognition, *Seventh International Conference on Working with Computer*, Kuala Lumpur, 2004 ,

[6] Zi-Quan Hong, Algebraic Feature Extraction of Image for Recognition *IEEE Transactions on Pattern Recognition* Vol. 24 No. 3 pp 211-219, 1999.

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

[8] W. Keith Nicholson*, Elementary Linear Algebra*. McGraw-Hill International Edition, Singapore, 2001

[9] Li Min Fu*, Neural Networks in Computer Intelligence,* McGraw- Hill International Editions, Singapore 1994.

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

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