

STEREO VISION SYSTEM FOR A BIN PICKING ADEPT ROBOT

*Hema Chengalvarayan Radhakrishnamurthy, Paulraj Murugesapandian,
Nagarajan Ramachandran, Sazali Yaacob*

School of Mechatronic Engineering
University Malaysia Perlis, 02600, Jejawi, Perlis, Malaysia.
Email: {hema, paul, nagarajan, s.yaacob}@unimap.edu.my

ABSTRACT

In bin picking applications, robots are required to pick up an object from a pile of stacked or scattered objects placed in a bin. To perform such tasks, identification of the objects to be picked using a vision system is indispensable. In this paper, a stereo vision based automated bin picking system is proposed which identifies the topmost object from a pile of occluded objects and computes its location. The proposed bin picking process consists of two modules namely object segmentation module and object localization module. In the segmentation module, an 'Acclimatized Top Object Threshold' [ATOT] algorithm is proposed for segmentation of topmost object and in the localization module, the location of the object is estimated by computing the 'x', 'y', 'z' co-ordinates of the object midpoint using a unified stereo imaging algorithm. The validity of the algorithms is experimentally verified for object pick and place operations using the object location co-ordinates. The developed stereo vision system was implemented and validated for bin pick and place operations on an Adept Cobra 600 Robot.

Keywords: *Stereo Vision, Segmentation, Stereo Image Processing, Neural Networks, Bin Picking Robots.*

1.0 INTRODUCTION

Industrial robots used for assembly operations as well as pick and place operations can handle only objects placed in preprogrammed locations and orientation. The blind robot hence is used only for repetitive tasks; this design increases the cost of automation as multiple robots and expensive object feeding conveyor systems are required.

An obvious solution to this situation can be provided by depicting the manual assembly environment which has objects placed in different bins around the work area and required objects can be picked for assembly. Automation of this process requires an intelligent robot capable of handling objects in bins. Bin picking involves picking an object from a bin consisting of similar objects, the objects in the bin are generally jumbled and occlude each other; this poses a great challenge to the vision sensors in identifying the objects to be picked. Many researches on vision based bin picking have been limited to recognition of objects due to the complexity involved in segmenting the bin images. In recent years, much work have been reported on recognition of objects for the bin picking problem. Some concepts are limited to recognition of the objects while others concentrate on picking an object and analyzing the object for pose determination and recognition using database of the object images [1], [2], [3]. Stereo based approaches have been reported by some researchers; Rahardja and Kosaka [4] have developed an effective stereo based algorithm to find simple visual clues of complex objects, however the system requires human intervention for grasping tasks. Martin and et al [5] use stereo and CAD models for flexible assembly of industrial parts, stereo matching and structured light are used for the picking process while CAD models are used to determine the pose of the objects. An inspection module is presented to ensure correct mounting of the part.

Image segmentation is another well researched area which involves division of the image into meaningful sub regions and is an indispensable step in image analysis. If all the objects 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. One approach for solving the problem of identifying objects is by detecting and measuring all segments belonging to the object border and all internal angles [6]. Similarly, regional growth is a technique that starts at the known pixel points and extends to all neighbouring pixels that are similar in gray level, color 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 [7].

In contrast to the approaches mentioned above, this research attempts to provide a solution to segmentation and location of partially occluded bin objects using a stereo vision based approach. The objective of this research is to develop a vision system for an adept cobra robot to perform automated bin picking operations. This paper mainly concentrates on solving the segmentation and localization problem of partially occluded objects. The task of bin picking is split into two subtasks namely object segmentation and object location. The following sections discuss the two subtasks in detail. The stereo vision system methodology is detailed in Section 2 while the proposed segmentation algorithm using binary thresholding and image histogram is explained in Section 2.1. Section 2.2 elaborates on the object localization process and Section 2.3 describes the network architecture for computing the 'z' coordinate. Section 3 gives the experimental results obtained from the algorithms and the real time bin picking experiments conducted on an adept robot and finally section 4 presents the discussion and conclusion sections.

2.0 METHODOLOGY

The hardware of the stereo vision system consists of two Pulnix TM 6702 CCD machine vision cameras placed in a stereo rig. A base length of 70 mm is experimentally determined for the current application. The stereo rig is fixed to the arm of the robot. A vacuum gripper [Bellows style, 18 mm ϕ] is used for picking the bin objects. The size of the bin is 300 mm x 200 mm x 200 mm; however the effective area for stereo imaging is 260mm x 200 mm. Fig. 1. shows the vision system interfaced to the Adept Robot. Direct lighting of the bin is avoided to reduce brightness and albedo effects [8]. In the experimental process a bin consisting of partially occluding objects are chosen. Stereo images from the left and right cameras are acquired simultaneously using custom designed interface software. Stereo images acquired from the vision sensors are monochromatic images of size 640 x 480 pixels. The acquired images are initially pre-processed to improve the efficiency of the segmentation process. In the pre-processing stage the images are first resized to minimize the processing time and to improve the efficiency of the system without significant loss of information. Resized images of 128 x 96 pixels are experimentally found to be suitable.

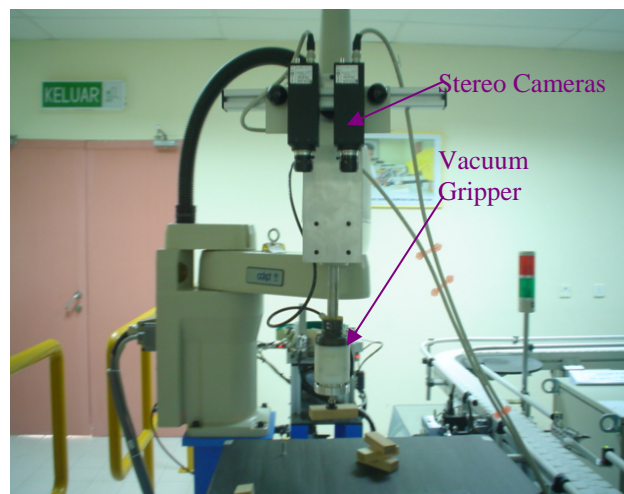


Fig. 1: Stereo Vision System Hardware

The topmost object in the bin is free of occlusions and is thus chosen as the desired object for pickup due to occlusions intensity levels of objects vary and the topmost is presumed to have the highest intensity levels in comparison to the rest of the objects. Image enhancement is done using a regional filter; which smoothens out the intensity of the image around the objects. The filtered image is then subjected to segmenting techniques.

2.1 Object Segmentation Module

The object segmentation module identifies the top most object from the cluster of objects in the bin. Since all the objects are partially occluded except the topmost object, separating the topmost object can be done using the binary segmentation technique. The topmost object can be segmented by applying a suitable threshold, which is commonly derived by trial and error. However, in real time automation applications, automatic detection of the threshold value is essential. In this research, we propose an algorithm for automatic detection of a global threshold which segments

the topmost object. The proposed algorithm is also compared with the more popular Otsu method [9]. The proposed algorithm determines the threshold from the histogram of the bin images to segment the topmost object.

The proposed ATOT algorithm is outlined as follows:

Step 1: Computed the histogram for the left and right gray scale images.

$counts_left(i), i=1, 2, 3, \dots, 256$

contains the number of pixels with a gray scale value of $(i-1)$ pixels for the left image.

$counts_right(i), i=1, 2, 3, \dots, 256$

contains the number of pixels with a gray scale value of $(i-1)$ pixels for the right image.

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

If $count_left(i) \neq 0$

$$t_left(i) = \log(count_left(i)) * (i-1) \quad (1)$$

If $count_right(i) \neq 0$

$$t_right(i) = \log(count_right(i)) * (i-1) \quad (2)$$

where $i = 1, 2, 3, \dots, 255$

Step 3: Compute the mean of all the logarithmic weighted gray scale for left and right images

$$tm_left = \frac{1}{256} \sum_{i=1}^{256} t_left(i) \quad (3)$$

$$tm_right = \frac{1}{256} \sum_{i=1}^{256} t_right(i) \quad (4)$$

Step 4: The threshold T is the minimum value of ' tm_left ' and ' tm_right '.

$$T = \min(tm_left, tm_right).$$

Fig. 2 shows the bin images before and after segmentation. Threshold of both the stereo images are computed separately and the minimum value of the two thresholds is applied as the threshold to both the images.

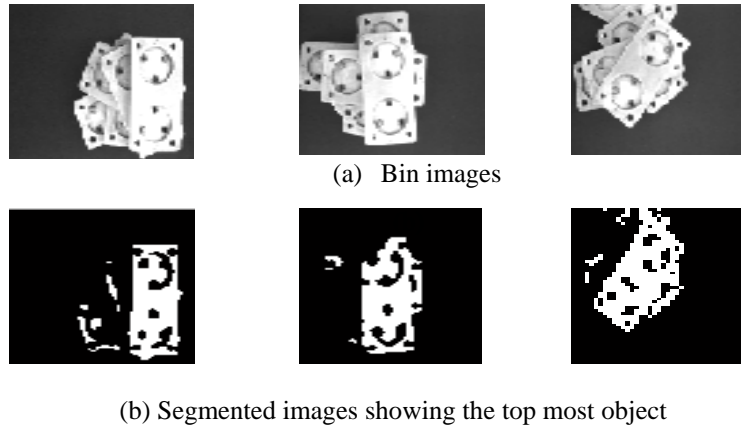


Fig. 2: Bin Image Segmentation

The gray images are converted to binary images by applying the threshold derived from the ATOT algorithm. The segmentation results of the proposed algorithm are compared with the more popular Otsu global threshold method. Fig. 3 shows the comparison between Otsu method and the ATOT method.

It is observed from Fig. 3 that the proposed algorithm gives better segmentation results than the Otsu method for segmentation of the topmost object in a bin image.

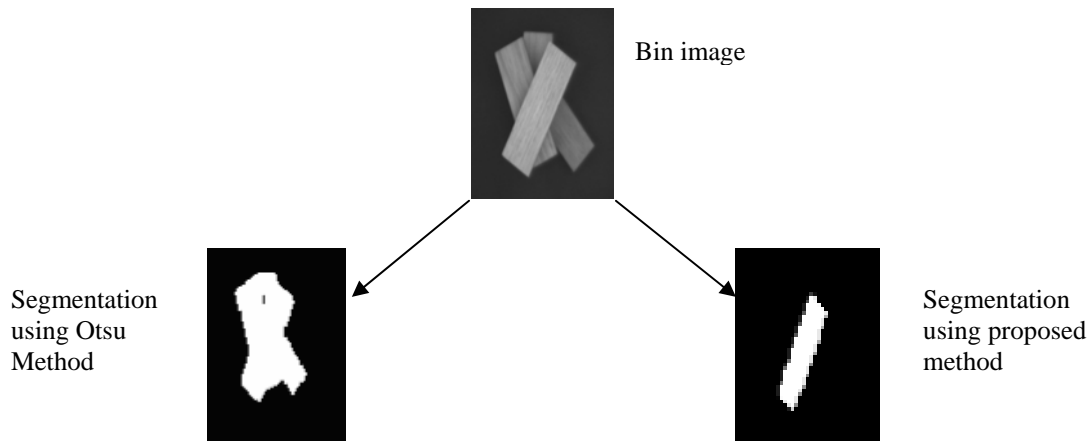


Fig. 3: Comparison results of Global Threshold Algorithms

2.2 Object Localization

The next phase is the computation of the location of the object with respect to the bin area. It is assumed that the midpoint of the object will be ideal to pick the object using the vacuum gripper. Hence the x , y and z co-ordinates of this point are to be provided to the robot. The ' x ' and ' y ' co-ordinates are computed by finding the centroid of the object image, while the z co-ordinate, which is the distance of the object from the sensors, is computed using stereo techniques and neural networks. An algorithm called the 'unified stereo imaging' is proposed to compute ' z ' which uses the distance information obtained from adding the stereo images to train a neural network.

The proposed unified stereo imaging algorithm is described in the following steps.

1. Object stereo images of the segmentation module are morphologically added.
2. Extract edge image of the added image using a canny edge detector.
3. Compute singular features of the edge image
4. Train a neural network with object features and distance [z] values for z co-ordinate computation.

Fig. 4 shows the flow diagram of the z co-ordinate estimation process. Object features are extracted from the edge image using singular value decomposition [SVD] of the image matrix. 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 [10]. Any ' $m \times n$ ' matrix A ($m \geq n$) can be written as the product of an ' $m \times m$ ' 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 [11]. 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 are algebraic features of an image, which represents the intrinsic attributes of the unified image [10].

Singular features of the topmost objects are extracted using the unified stereo imaging algorithm. The singular values obtained from the added images are fed as input to a feed forward neural network and the distance of the object is given as output. The network is trained with 80% data samples. The trained network is tested with real time images to estimate the object distance.

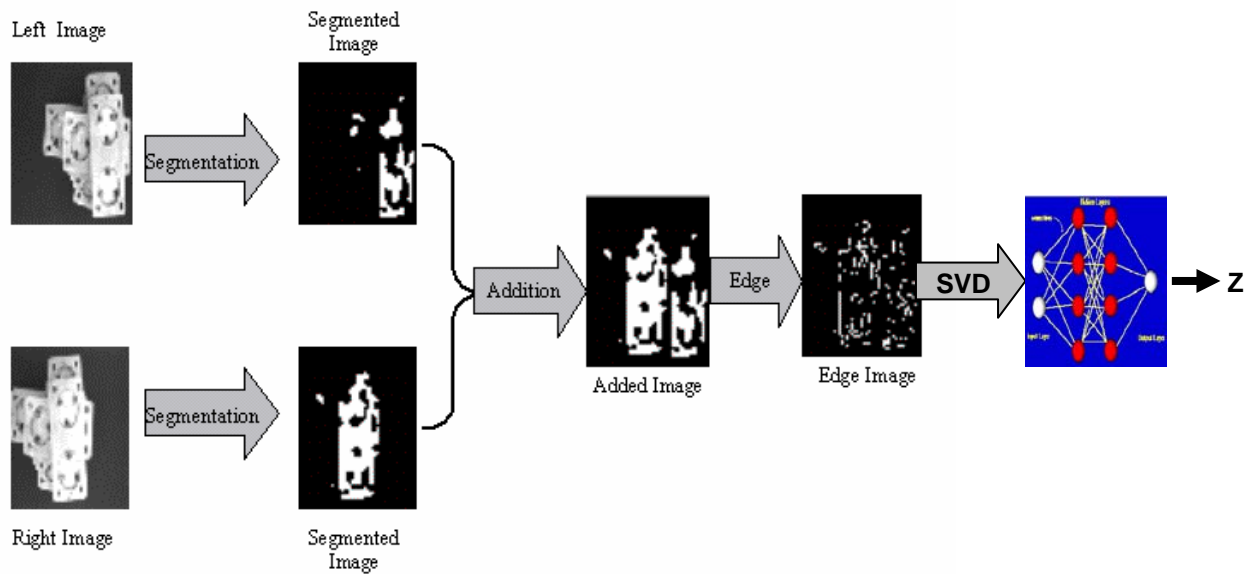


Fig. 4: Flow Diagram for the estimation of Z co-ordinate

3.0 NEURAL NETWORK ARCHITECTURE

The Feed Forward Neural Network architecture consists of three layers, the input layer, the hidden layer and the output layer. Ten significant singular values are chosen and the values close to zero or zero are not considered as they do not contribute to the data set [10]. These 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. The initial weights are normalized using equation (5) [12].

$$w_{1j}(new) = \frac{w_{1j}(old)}{\sqrt{w_{1j}^2 + w_{2j}^2 + \dots + w_{nj}^2}} \quad (5)$$

where $j = 1, 2, 3, \dots, p$
 n is the number of input units
 p is the 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 [13]. The cumulative error versus epoch plot of the trained neural network is shown in Fig. 3. The cumulative error is the sum squared error for each epoch and is given by:-

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

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.

4.0 EXPERIMENTAL RESULTS

In the experimental study six wooden blocks were used as objects. The objects are placed in the bin, partially occluding each other. About 53 stereo image of the bin for various positions of the six objects are acquired. The stereo images are first processed using the object segmentation module and object location module to extract the feature data of the topmost object. In the training phase, the neural network is trained using 43 sample feature data, the network is trained using the back propagation algorithm. In the testing phase, the network is tested with 53 sample data. In the offline testing process, the proposed network is found to successfully compute the distance or z co-ordinate for 49 images with a success rate of 93.87%. Fig. 5 shows the cumulative error versus epoch plot of the neural network. Table I shows the training parameters and test results of the neural network.

In the online real-time bin picking, the stereo vision system is interfaced with an Adept Cobra Robot as shown in Fig. 1. The vision software is interfaced with the Adept Robot to perform bin picking operations. In the experimental phase the vision system segments the topmost object and computes the midpoint [x, y, and z co-ordinates] of the topmost object. The vacuum gripper is programmed to move to the computed location to pick the object and place the object in desired location outside the bin. The process is repeated for all objects in the bin. The robot successfully picked 5 objects from a bin containing 6 partially occluded objects.

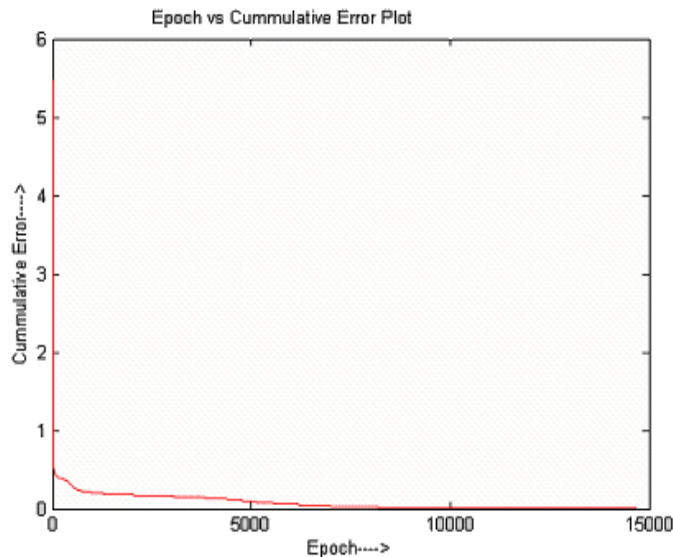


Fig. 5: Epoch versus cumulative error plot of neural network

Table 1: Training Parameters and Results of Neural Network

| Sl. No. | Parameters | Value |
|---------|--------------------|------------|
| 1 | Input Neurons | 10 |
| 2 | Hidden Neurons | 18 |
| 3 | Output Neurons | 1 |
| 4 | Learning rate | 0.1 |
| 5 | Momentum Factor | 0.3 |
| 6 | Training samples | 43 |
| 7 | Training Tolerance | 0.0005 |
| 8 | No. of Epoch | 14719 |
| 9 | Testing tolerance | 0.01 |
| 10 | Testing Samples | 49 |
| 11 | Training Time | 30 seconds |
| 12 | Success Rate | 93.87% |

5.0 CONCLUSION

A stereo vision system for segmentation and object location computation in bin picking applications is presented. Two algorithms namely ATOT and Unified Stereo Imaging algorithms are proposed. The ATOT algorithm is found to be robust in the absence of albedo effects; however for images with over illumination, the segmentation results were unsatisfactory. The unified stereo imaging algorithm is capable of computing accurate values for 'x' and 'y', while the 'z' value is computed with an average error of ± 1 mm; however pick up operations were unaffected due to the bellow style vacuum gripper which has a tolerance of up to ± 2 mm. In automated bin picking experimentation, it was observed that efficiency of the pick and place operation entirely depended on the segmentation process and the segmentation efficiency depended on the illumination levels of the objects. The main constraint of the vision system is in providing optimal lighting of the bin. The real time experimental results prove the feasibility of the proposed algorithms for bin picking operations. Future research efforts will be directed towards improving the segmentation results and in minimizing the error in computing the 'z' co-ordinate values.

REFERENCES

- [1] Ayako Takenouchi, Naoyoshi Kanamaru & Makoto Mizukawa, "Hough-space-based Object Recognition Tightly Coupled with Path Planning for Robust and Fast Bin-picking" in *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," in *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] Krisnawan Rahardja & Akio Kosaka, "Vision-Based Bin –Picking : Recognition and Localization of Multiple Complex Objects Using Simple Visual Cues" in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 1996, Vol. 3, pp. 1448 – 1457.
- [5] B. Martin, B. Bachler & S. Stefan, "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*. USA. 2000.
- [6] L. Paulo & O. Eugenio,"Identification of Partially Occluded Objects Using Object Chain Code" in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems Yokohama*, Japan, July 1993, pp. 1829-1834.
- [7] P. W. M. Tsang & P. C. Yuen, "Recognition of Partially Occluded Objects", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 23, No.1, January/February 1993.
- [8] D. Arcangelo, A. Nicola, A. Giovanni, C. Laura, C. Maria & S. Ettore, "A Model-Based 3-D Vision System for Bin-Picking", *IEEE Transactions on Circuits and Systems*, Vol. 35 No.5, May 1988.
- [9] N. Otsu, "A threshold selection method from gray level histograms", *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-9, 1979, No. 1, pp. 62 – 66.
- [10] H. Zi-Quan, "Algebraic Feature Extraction of Image for Recognition", *IEEE Transactions on Pattern Recognition*, Vol. 24 No. 3, 1991, pp. 211-219.
- [11] S. W. David, *Fundamentals of Matrix Computations*, John Wiley & Sons, 2001.
- [12] S. N. Sivanandam & M. Paulraj, *Introduction to Artificial Neural Networks*, Vikas Publishing House, India, 2003.
- [13] F. Laurene, *Fundamentals of Neural Network Architecture and Applications*, Prentice Hall, USA, 1990.

BIOGRAPHY

Hema Chengalvarayan Radhakrishnamurthy obtained her Master of Science (Electrical and Electronics Engineering) from Universiti Malaysia Sabah in 2005. Currently, she is a lecturer at the School of Mechatronic Engineering, University Malaysia Perlis. Her research areas include Robot and Machine Vision and Biomedical Signal Processing. She has published a number of papers related to these areas. She is also a member of IEEE.

Paulraj Murugesapandian received his PhD degree in Computer Science and Engineering from Bharathiyar University, India in 2000. He is currently working as an Associate Professor at the School of Mechatronic Engineering, University Malaysia Perlis. His research interests include Design and Development of Intelligent Algorithms, Fuzzy Systems, Neural Networks, Acoustics, Image processing and Signal Processing. He has authored a book on Neural Networks and has published more than 100 papers in related fields.

Nagarajan Ramachandran is a Professor at the School of Mechtronic Engineering, Universiti Malaysia Perlis. His research areas include Robotics, Vision Systems and Control Systems. He is a Senior Member of IEEE. He has published more than 350 papers related to these areas. Many of his papers have received best paper awards in journals and conferences.

Sazali Yaacob is a Professor at the School of Mechatronic Engineering, Universiti Malaysia Perlis. His research areas include Control Systems, Robotics and Vision Systems and have published more than 100 papers in these areas.