# Zernike Based Scene Categorization using Support Vector Machines

Devendran. V[1], Hemalatha Thiagarajan[2], Amitabh Wahi[3]

**Abstract**

Natural scene categorization is a fundamental process of human vision that allows us to efficiently and rapidly analyze our surroundings. Thousands of images are generated every day, which implies the necessity to classify, organize and access them using an easy, faster and efficient way. Scene categorization, the classification of images into semantic categories (e.g., coast, mountains, highways and streets) is a challenging and important problem nowadays. Many different approaches concerning scene categorization have been proposed in the last few years. This paper presents a different approach to classify 'MIT-Street' and 'MIT-highway' scene categories using Zernike moments and Support Vector Machines. Radial Basis Function with p1=5 is used for the paradigm of Support Vector Machines. This work is implemented using various blocking models and the comparative results are proving the importance of blocking and the efficiency of classifier towards scene classification problems. This complete work is carried out using real world data set.

**Keywords:** Scene Categorization, Support Vector Machine, Zernike Moments.

[1,3] Dept.of Computer Applications & Information Technology Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India. E-mail : svdevendran@yahoo.com, awahi@rediffmail.com

[2] Dept. of mathematics, National Institute of Technology, Trichy, Tamilnadu,India. Email: hema@nitt.edu.

## 1. Introduction

Understanding the robustness and rapidness of human scene categorization has been a focus of investigation in the cognitive sciences over the last decades [1][2][3]. At the same time, progress in the area of image understanding has prompted computer vision researchers to design computational systems that are capable of automatic scene categorization. Classification is one of several primary categories of machine learning problems [4]. Florica Mindru and et al., [5] gives a systematic overview of moment invariants for several combinations of deformations and photometric changes. Papers [6], [7] and [8] give very promising results in the classification of indoor-outdoor scene image and manmade-natural classification. Ian Stefan Martin [9] presents in his doctoral work, the techniques for robust learning and segmentation in scene understanding. Zernike Moments [10] are important shape descriptors in computer vision. There are two types of shape descriptors: contour-based shape descriptors and region-based shape descriptors. Manuele Bicego et al. [11] give a new approach to scene analysis under unsupervised circumstances. In [12][13], Bosch et al. present a scene description and segmentation system capable of recognizing natural objects (e.g., sky, trees, grass) under different outdoor conditions. In this study, a computer vision system recognizing objects in captured images is established using Zernike Moments (ZM). The organization of the paper is as follows: section II deals with Support Vector Machines, section III gives Zernike feature extraction method, section IV deals with section VI gives concluding remarks on experiment results.

## 2. Support Vector Machines

Support vector machine is a relatively new pattern classifier introduced by Vapnik [14]. A SVM classifies an input vector into one of two classes, with a decision boundary developed based on the concept of structural risk minimization (of classification error) using the statistical learning theory. The SVM learning algorithm directly seeks a separating hyperplane that is optimal by being a maximal margin classifier with respect to training data. For non-linearly separable data, the SVM uses kernel method to transform the original input space, where the data is non-linearly separable, into a higher dimensional feature space where an optimal linear separating hyperplane is constructed. On the basis of its learning approach, the SVM is believed to have good classification rate for high-dimensional data. Consider the problem of image classification where X is an input vector with 'n' dimensions. The SVM performs the following operation involving a vector $W = (w_1, ..., w_n)$ and scalar b:

$$f(X) = \text{sgn}(W \bullet X + b) \quad (1)$$

'MIT-street' category is recognized if the sign of f(X) is Positive otherwise '*MIT-highways*' category is assumed. Consider a set of training data with $l$ data points from two classes. Each data is denoted by $(X_i, Y_i)$, where i=1, 2,..., $X_i = (x_{i1}, ..., x_{in})$, and $Y_i \in \{+1, -1\}$. Note that $Y_i$ is a binary value representing the two classes. The task of SVM learning algorithm is to find an optimal hyperplane (defined by W and b) that separates the two classes of data. The hyperplane is defined by the equation:

$$W \bullet X + b = 0 \quad (2)$$

Where X is the input vector, W is the vector perpendicular to the hyperplane, and b is a constant. The graphical representation for a simple case of two-dimensional input (n = 2) is illustrated in Fig. 1. According to this hyperplane, all the training data must satisfy the following constraints:

$$W \bullet X_i + b \geq +1 \text{ for } \forall_i = +1$$
$$W \bullet X_i + b \geq -1 \text{ for } \forall_i = -1 \quad (3)$$

which is equivalent to :

$$y_i(W \bullet X_i + b) \geq 1 \ \forall_i = 1, 2, ...., l \quad (4)$$

There are many possible hyperplanes that separate the training data into two classes. However, the optimal separating hyperplane is the unique one that not only separates the data without error, but also maximizes the

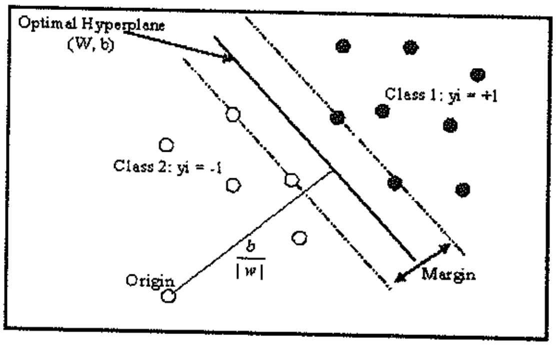



**Figure 1:Optimal Separating Hyperplane for 2-Dimensional Two-Class Problem**

margin, i.e., maximizes the distance between the closest vectors in both classes to the hyperplane [14] . As shown in Fig. 1, the margin, *P*, is the sum of the absolute distance between the hyperplane and the closest data points in each class. It is given by:

$$\rho = \min \frac{|W.X_i + b|}{\|W\|} + \min \frac{|W.X_i + b|}{\|W\|} = \frac{2}{\|W\|} \quad (5)$$

Here, the first min is over $X_i$ of one class and the second min is over $X_i$ of the other class. Therefore, the optimal separating hyperplane is the one that maximizes 2/||W||, subject to constraints (4). It is mathematically more convenient to replace maximization of 2 / ||W|| with the equivalent minimization of $\|W\|^2/2$ subject to constraints

(4), which can be solved by the Lagrangian formulation:

$$\min\ L = \frac{1}{2}\|W\|^2 - \sum_{i=1}^{l}\alpha_i\left[y_i(W.X_i + b) - 1\right] \quad (6)$$

where $\alpha_i$ is the Lagrange multiplier $\alpha_i$ (>=0, i=1,2,..$l$). The Lagrangian has to be minimized with respect to W and b, and maximized with respect to $\alpha_i$. The minimum of the Lagrangian with respect to W and b is given by:

$$\frac{\partial L}{\partial W} = 0 \Rightarrow W = \sum_{i=1}^{l}\alpha_i X_i y_i \quad (7)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l}\alpha_i y_i = 0 \quad (8)$$

Substituting (7) and (8) into (6), the primal minimization problem is transformed into its dual optimization problem of maximizing the dual Lagrangian $L_D$ with respect to :

$$\max\ L_D = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (X_i \bullet X_j) \quad (9)$$

subject to

$$\sum_{i=1}^{l}\alpha_i y_i = 0 \quad (10)$$

$$\alpha_i \geq 0 \ \forall\ i=1,\dots,l \quad (11)$$

Thus, the optimal separating hyperplane is constructed by solving the above quadric programming problem defined by (9)-(11). In this solution, those points have non-zero Lagrangian multipliers ($\alpha_i$ > 0) are termed support vectors. Support vectors satisfy the equality in the constraint (4) and lie closest to the decision boundary (they are circles in Fig. 1, lying on the dotted lines on either side of the separating hyperplane). Consequently, the optimal hyperplane is only determined by the support vectors in the training data. Based on the $\alpha_i$ values obtained, W can be calculated from (7). b can be obtained by using the Karush-Kuhn-Tucker(KKT) complementary condition for the primal Lagrangian optimization problem:

$$\alpha_i\left[y_i(W.X_i + b) - 1\right] = 0 \quad \forall_i = 1,..,l \quad (12)$$

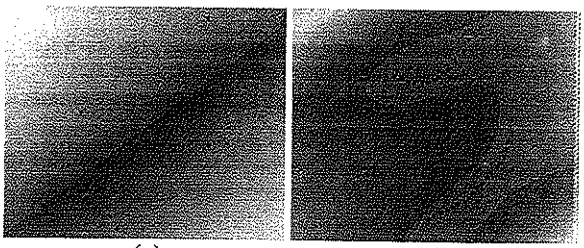
One b value may be obtained for every support vector (with $\alpha_i > 0$). Burges [15] recommends that the average value of b be used in the classification. With this solution, the SVM classifier becomes

$$f(X) = \mathrm{sgn}(W \bullet X + b) = \mathrm{sgn}\Big(\sum_{\forall i, \alpha_i > 0} y_i\alpha_i(X_i \bullet X) + b\Big) \quad (13)$$

(13) Note that, in (13), one only needs to make use of $X_i$, $Y_i$ and $\alpha_i$ of the support vectors, while X is the input vector to be classified. When a linear boundary is inappropriate (i.e., no hyperplane exists to separate the two classes of data), the extension of above method to a more complex decision boundary is accomplished by mapping the input vectors $X \in R^n$ into a higher dimensional feature space H through a non-linear function $\phi : R^n \rightarrow H$. In H, an optimal separating hyperplane is then constructed using training data in the form of dot products $\phi(X_i) \bullet \phi(X_j)$ instead of the $X_i \bullet X_j$ term in (9). To avoid the expensive computations of $\phi(X_i) \bullet \phi(X_j)$ in the feature space, it is simpler to employ a kernel function such that

$$K(X_i.X_j) = \phi(X_i) \bullet \phi(X_j) \quad (14)$$

Thus, only the kernel function is used in the training algorithm, and one does not need to know the explicit form of $\phi$. The computation in (15) results in some restrictions on the form and parameter values of non-linear functions that can be used as the kernel functions. Detailed discussions can be found in [14] and [15]. Some commonly used kernel functions are:



(a) (b)

**Figure 2: Representation of (a) Linearly Separable (b) Non-Linearly Separable**

Polynomial function: $K(X_i, X_j) = (X_i.X_j + 1)^d$ (15)

Radial basis function: $K(X_i, X_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$ (16)

Sigmoid function: $K(X_i, X_j) = \frac{1}{1 + e^{[v(X_i.X_j) - \delta]}}$ (17)

Where d is a positive integer, and, $\sigma$, $v$ and $\delta$ are real constants. These four parameters must be defined by the user prior to SVM training. With the use of a kernel function, the SVM is capable of performing non-linear classification of input X. Then, (13) becomes,

$$f(X) = \operatorname{sgn}\left( \sum_{\forall_i, \alpha_i > 0} y_i \alpha_i K(X_i, X) + b \right) \quad (18)$$

The hyperplane and support vectors used to separate the linearly separable data are shown in Fig. 2(a). And the hyperplane and support vectors used to separate the non-linearly separable data are shown in Fig. 2(b). Radial basis kernel function with p1=5 used for this non-linear classification. Individual colors represents particular each class of data.

### 3. Zernike Moments

Zernike polynomials were first proposed in 1974 by Zernike. Their moment formulation appears to be one of the most popular, outperforming the alternatives (in terms of noise resilience, information redundancy and reconstruction capability). Moments have been widely used in image processing applications through the years. Geometrical, central and normalized moments were for many decades the only family of applied moments. The main disadvantages of these descriptors were their disability to fully describe an object in a way that, using the moments set, the reconstruction of the object could be possible. In other words they weren't orthogonal. The kernel of Zernike moments is a set of orthogonal Zernike polynomials defined over the polar coordinate space inside a unit circle. The two-dimensional Zernike moments of order p with repetition q of an image intensity function f(r, θ) are defined as;

$$Z_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 V_{pq}(r,\theta) f(r,\theta) r\,dr\,d\theta; \; |r| \le 1 \quad (19)$$

where Zernike polynomials Vpq(r, θ) are defined as:

$$V_{pq}(r,\theta) = R_{pq}(r) e^{-\hat{j}q\theta}; \; \hat{j} = \sqrt{-1} \quad (20)$$

and the real-valued radial polynomials, Rpq(r), is defined as follows:

$$R_{pq}(r) = \sum_{k=0}^{\frac{p-|q|}{2}} (-1)^k \frac{(p-k)!}{k!\left(\frac{p+|q|}{2} - k\right)!\left(\frac{p-|q|}{2} - k\right)!} r^{p-2k} \quad (21)$$

where $0 \le |q| \le p$ and $p - |q|$ is even.

If N is the number of pixels along each axis of the image, then the discrete approximation of equation (19) is given as:

$$Z_{pq} = \lambda(p,N) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{pq}(r_{ij}) e^{-\hat{j}q\theta_{ij}} f(i,j); 0 \le r_{ij} \le 1 \quad (22)$$

where λ(p, N) is normalizing constant and image coordinate transformation to the interior of the unit circle is given by

$$r_{ij} = \sqrt{x_i^2 + x_j^2}; \theta = \tan^{-1}\left(\frac{y_i}{x_i}\right); x_i = c_1 i + c_2;$$
$$y_i = c_1 j + c_2; \quad (23)$$

Since it is easier to work with real functions, Zpq is often split into its real and imaginary parts, $Z^c_{pq}, Z^s_{pq}$ is given below:

$$Z^c_{pq} = \frac{2(p+1)}{\pi} \int_0^{2\pi} \int_0^1 R_{pq}(r) \cos(q\theta) f(r,\theta) r\,dr\,d\theta \quad (24)$$

$$Z^s_{pq} = \frac{2(p+1)}{\pi} \int_0^{2\pi} \int_0^1 R_{pq}(r) \sin(q\theta) f(r,\theta) r\,dr\,d\theta \quad (25)$$

where $p \ge 0$, $q > 0$.

For, the implementation, square image (NxN) is transformed and normalized over a unit circle; i.e., $x^2 + y^2 \leq 1$, in which the transformed unit circle image is bounding the square image. In this transformation,

$$\lambda(p,N) = \frac{4(p+1)}{(N-1)^2\Pi}; c1 = \frac{\sqrt{2}}{N-1}; c2 = \frac{-1}{\sqrt{2}} \quad (26)$$

Therefore,

$$x_i = \frac{\sqrt{2}}{N-1}i + \frac{-1}{\sqrt{2}} \text{ and } y_j = \frac{\sqrt{2}}{N-1}j + \frac{-1}{\sqrt{2}} \quad (27)$$

## 4. Proposed Work

In classification, a system is trained to recognize a type of example or differentiate between examples that fall in separate categories. In order to successfully accomplish this, the classifier must have sufficient prior knowledge about the appearance of the object. This paper is trying to recognize the scenes of two different categories called '*MIT-street*' and '*MIT-highways*'. This work is implemented and compared using three kinds of blocking models, i.e., Block-1, Block-4 and Block-16. Block-1 considers the image as it is, Block-4 divides the image into four equal parts and Block-16 divides the image into sixteen equal parts. Classification is carried out in all the three kinds of blocking and the results are then compared. The detailed description of our proposed work is shown in Fig. 3.

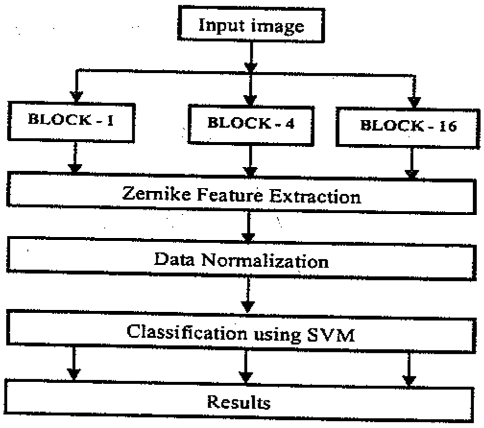



**Figure 3 : Detailed Description of Proposed Work**

In Block-I method, raw input images are received, as it is, for the classification of scenes. Then, Zernike features are extracted from the raw images and normalized using zero-mean normalization method in order to make the data within certain range to improve the classifiers performance. Normalized features are then given for the classification in support vector machines using radial basis kernel function. 200 samples are used for training and another 200 samples are used for testing of scene categorization problem. These samples are taken such a manner that there is no overlapping between training and testing samples. The results are then discussed in terms of true positive, true negative, false positive and false negative. The same sequences are followed in other two blocking methods i.e., Block-4 and Block-16, considering that Block-4 divides the input image into four equal parts and Block-16 divides the image into 16 equal parts. Zernike feature extraction methods are applied on all the blocks of the image and kept combined as a feature set of an image. Then, the data is normalized, classified and the results are discussed in Table 1. Support Vector Machines with Radial Basis Kernel Function is used as classifier for this scene categorization problem. The sample images of 'MIT-highways' are given in Fig. 4 and the sample images of 'MIT-street' are given in Fig. 5. These sample scene category images are taken from the Ponce Research Group [16] which contains 15 different scene categories with 250 samples each.

## 5. Implementation

Using normalized zernike moment features, support vector machine is trained and tested to categorize the 100 samples from '*MIT-street*' and 100 samples from '*MIT-highways*'. In testing phase, the other set of 200 samples are used including 100 samples from '*MIT-street*' and 100 samples from 'MIT-highways'. The detailed implemented works are given in Table 1. True Positive (TP) is the number of positive samples correctly classified. True Negative (TN) is the number of positive samples wrongly classified. False Positive (FP) is the number of negative samples correctly classified. False Negative (FN) is the number of negative samples wrongly classified. Percentage of classification rate is measured by averaging the True Positive and False Positive values.

**Block-1:** The image is considered as a single block, feature vector of 36x1 is extracted and normalized. Support Vector Machine classifies the data and gives true positive of 90% and false positive of 74%. Hence, the classification rate of 82.0 % is achieved in 716.98 seconds.

**Block-4:** The input image is divided into four blocks; feature vector of 4x36=144 is extracted and normalized. Support Vector Machine classifies the data and gives true positive of 93% and false positive of 78%. Hence, classification rate is of 85.5% is achieved in 656.16 seconds.

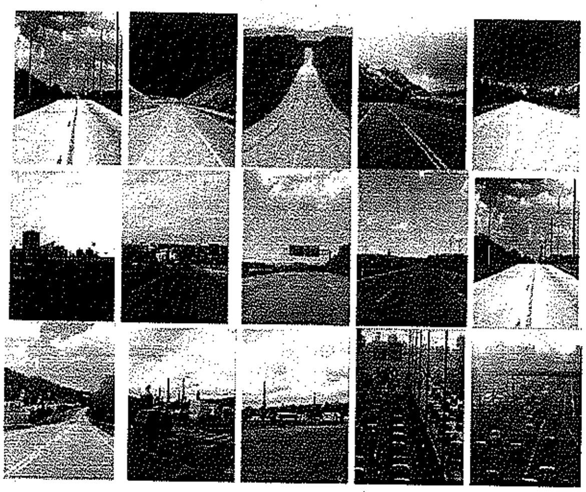


**Figure 4: Sample Images of 'MIT-highway' Category**

**Block-16:** The input image is divided into sixteen blocks; feature vector of 16x36=576 is extracted and normalized. Support Vector Machine classifies the data and gives true positive of 85% and false positive of 100%. Hence, classification rate is of 92.5% is achieved in 674.16 seconds.

**Figure 5: Sample images of 'MIT-street" category**

**Table 1: Performance Of Classification Of Support Vector Machines**

| Experiment Type | TP | TN | FP | FN | Classify % | *Execution Time (in Sec)* |
|---|---|---|---|---|---|---|
| Block – 1 | 90 | 10 | 74 | 20 | 82 | 716.98 |
| Block – 4 | 93 | 07 | 78 | 22 | 85.5 | 656.16 |
| Block - 16 | 85 | 15 | 100 | 0 | 92.5 | 674.16 |

## 6. Conclusion

This paper concentrates on the categorization of images as '*MIT-street*' scenes or '*MIT-highways*' scenes. This paper discusses Zernike based scene classification problem using Support Vector Machines. Sample Image Database is taken from Ponce Research Group [16]. The performance of the classifier is given in Table 1. This shows that the Support Vector Machines with Radial Basis Kernel Function is giving its best of 92.5% classification rate with 674.16 seconds when the images are divided into sixteen equal blocks. This also proves the importance of blocking oriented approach towards the classification problems.

This work can be further extended to classify other natural scene categories [16] using various kernel functions. This complete work is implemented using SVM Toolbox in Matlab 6.5.

## References


[1] A. Guerin-Dugue and A. Oliva, *"Classification of scene photographs from local orientations features"*, Pattern Recognition Letters 21, PP. 1135-1140, 2000.

[2] A. Chella, M. Frixione and S. Gaglio, *"Understanding dynamic scenes"*, Artificial Intelligence 123, PP. 89-132, 2000.

[3] Ilkka Autio and Tapio Elomaa, *"Flexible view recognition for indoor navigation based on Gabor filter and support vector machines"*, Pattern Recognition 36, PP. 2769-2779, 2003.

[4] Navid Serrano, Andreas E. Savakis and Jiebo Luo, *"Improved scene classification using efficient low-level features and semantic cues"*, Pattern Recognition 37, PP. 1773-1784,2004.

[5] Florica Mindru and et al., *"Moment invariants for recognition under changing viewpoint and illumination"*, Computer Vision and Image understanding 94, PP. 3-27, 2004.

[6] Matti Pietikainen, Tomi Nurmela, Topi Maenpaa, Markus Turtinen, *"View-based recognition of real-world textures"*, Pattern Recognition 37, PP. 313-323, 2004.

[7] Matthew Boutell and Jiebo Luo, *"Beyond pixels: Exploiting camera metadata for photo classification"*, Pattern Recognition 38, PP. 935-946, 2005.

[8] Andrew Payne and Sameer Singh, *"Indoor vs. outdoor scene classification in digital photographs"*, Pattern Recognition 38 , 1533-1545, 2005.

[9] Ian Stefan Martin, *"Robust Learning and Segmentation for scene Understanding"*, PhD Thesis, Dept. of Electrical Engineering and Computer science, Massachusetts Institute of Technology, May 2005.

[10] Mohamed Rizon, et al., *"Object Detection using Geometric Invariant Moment"*, American Journal of Applied Sciences 2(6): PP. 1876-1878, 2006.

[11] Manuele Bicego, Marco Cristani and Vittorio Murino, *"Unsupervised scene analysis: A hidden markov model approach"*, Computer vision and image understanding 102, PP. 22-41, 2006.

[12] A. Bosch, X. Munoz and J. Freixenet, *"Segmentation and description of natural outdoor scenes"*, Image and Vision computing 25, PP. 727-740, 2007.

[13] Anna Bosch, Xavier Munoz and Robert Marti, *"Which is the best way to organize/classify images by content?"*, Image and vision computing 25, PP. 778-791,2007.

[14] Vapnik .V.N, *"The support vector method of function estimation. In: Generalization in Neural Network and Machine Learning"*, Springer-Verlag, New York, NY, PP. 239-268, 1998.

[15] Burges. C.J.C, *"A tutorial on support vector machines for pattern recognition"*, data mining and knowledge discovery 2(2), PP. 121-167, 1998. Available from http://svm.first.gmd.de

[16] www-cvr.ai.uiuc.edu/ponce_grp/data.


***Author's Biography***

Devendran V received his Bachelor of Science in Mathematics from Vivekananda College, Madurai Kamaraj University, Madurai and Master of Computer Applications from Ayya Nadar Janaki Ammal College, Sivakasi in 1997 and 2000 respectively.

Currently he is working as an Assistant Professor in the Department of Computer Applications, Bannari Amman Institute of Technology, Sathyamangalam, Tamil nadu. Currently he is pursuing his Doctoral Programme in National Institute of Technology, Trichirappalli, Tamil nadu. He has worked as a co-investigator in a research project funded by DRDO, New Delhi during 2003 and 2005. He has published three research papers in refereed journals and four papers in IEEE International Conferences. His research interest includes Object Recognition, Pattern Recognition, Texture Analysis, Neural Networks and Support Vector Machines.

Hemalatha Thiagarajan received Doctor of Philosophy from University of Taxas, USA. She has over 30 years of academic and research experience in reputed educational institution. She is now working as a professor in Mathematics at National Institute of Technology, Trichirappalli, Tamil nadu. She has published over 20 papers in various national, international journals and conferences. Currently she is guiding five Ph. D. research scholars. Her area of specialization includes operations research, algorithms, neural networks and image processing.

Amitabh Wahi received his Doctor of Philosophy from Banaras Hindu University, Varanasi in 1999. He is currently working as a Professor in the Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam, Tamil nadu. Currently he is guiding three Research Scholars. He has acted as a Principal Investigator in a project funded by DRDO, New Delhi during 2003 and 2005. His area of specialization includes soft computing, pattern recognition, neural networks and image processing. He has published over 30 papers in various national, international conferences and journals.