Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 94298, 10 pages
doi:10.1155/2007/94298

Research Article

Indoor versus Outdoor Scene Classification Using Probabilistic Neural Network

Lalit Gupta, Vinod Pathangay, Arpita Patra, A. Dyana, and Sukhendu Das

Visualization and Perception Laboratory, Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai-600 036, India

Received 1 December 2005; Revised 22 May 2006; Accepted 27 May 2006

Recommended by Stefan Winkler

We propose a method for indoor versus outdoor scene classification using a probabilistic neural network (PNN). The scene is initially segmented (unsupervised) using fuzzy C-means clustering (FCM) and features based on color, texture, and shape are extracted from each of the image segments. The image is thus represented by a feature set, with a separate feature vector for each image segment. As the number of segments differs from one scene to another, the feature set representation of the scene is of varying dimension. Therefore a modified PNN is used for classifying the variable dimension feature sets. The proposed technique is evaluated on two databases: IITM-SCID2 (scene classification image database) and that used by Payne and Singh in 2005. The performance of different feature combinations is compared using the modified PNN.

1. INTRODUCTION

Classification of a scene as belonging to indoor or outdoor is a challenging problem in the field of pattern recognition. This is due to the extreme variability of the scene content and the difficulty in explicitly modeling scenes with indoor and outdoor content. Such a classification has applications in content-based image and video retrieval from archives, robot navigation, large-scale scene content generation and representation, generic scene recognition, and so forth. Humans classify scenes based on certain local features along with the context or association with other features. This context is learned by experience (training). Some examples of such local features are the presence of trees, water bodies, exterior of buildings, sky in an outdoor scene and the presence of straight lines or regular flat-shaded objects or regions such as walls, windows, artificial man-made objects in an indoor scene. Also, the types of features that humans perceive from images are based on color, texture, and shape of local regions or image segments. In this work, we represent the image as a collection of segments that can be of arbitrary shape. From each segment color, texture, and shape features are extracted. Therefore, the problem of indoor versus outdoor scene classification is a feature set classification problem where the number of feature vectors in the feature set is not constant, as the number of segments in an image varies. Also, there is no implicit ordering of the feature vectors in the feature set. This rules out the use of classifiers that take fixed dimension input feature vectors for classification. Hence we propose a modified probabilistic neural network that can handle variability in the feature set dimension.

The rest of this paper is organized as follows. The following section reviews existing work done in the indoor versus outdoor scene classification. Section 3 discusses the unsupervised segmentation of the scenes using fuzzy C-means clustering (FCM). The extraction of features from segments is described in Section 4. Section 5 describes PNN and its modification for scene classification. Section 6 discusses the results of the proposed technique on two databases. Section 7 concludes the paper and gives directions of future work.

2. REVIEW

The approaches used for scene classification (indoor versus outdoor) rely on features such as, edges, color, texture, and shape properties. Saber and Tekalp [1] integrated color, edge, shape, and texture features for region-based image annotation and retrieval. The classifiers used are Bayesian, independent component analysis (ICA), principal component analysis (PCA), and artificial neural network (ANN). Payne and Singh [2] had proposed a technique based on analyzing straightness of an edge in images. They classified images based on the hypothesis that indoor images have a greater proportion of straight edges compared to outdoor images. They used multiresolution estimates on edge straightness to improve the efficiency of the technique. Their method failed when images contain some objects prevalent in both indoor and outdoor environments. For 872 images they obtained 87.70% accuracy on gray-level image and 90.71% on subsampled image.

Jain and Vailaya [3] proposed an efficient retrieval of images from large databases exploiting important visual clues like color and shape content of an image. Experimental results on a database of 400 trademark images showed that integrated color- and shape-based feature provided 99% of the images being retrieved within the top two positions. Vailaya et al. [4] had shown that high-level classification problem (city images versus landscapes) can be solved from simple low-level features trained for the particular classes. They developed a procedure for measuring the saliency of a feature towards a classification problem based on intraclass and interclass distance distributions. The procedure is used to determine the discrimination power of the features: color histogram, color coherence vector, DCT coefficient, edge direction histogram, and edge direction coherence vector. Among them edge direction-based features had shown maximum discriminative power. For classification, a weighted *k-NN* had been used resulting in an accuracy of 93.9% when evaluated on an image database of 2216 images using leave-one-out strategy. Iqbal and Aggarwal [5] developed an approach for content-based image retrieval based on isotropic and anisotropic mappings. Isotropic mapping is invariant to the action of planar Euclidean group, translation, rotation, and reflection of image data and hence, invariant to orientation and position. Anisotropy mapping is variant to all these transformations. Isotropic mappings is represented by structure extraction via perceptual grouping and color histogram. The representation for anisotropic mapping is considered to be a channel energy model comprised of even-symmetric Gabor filters for texture analysis. They used 521 images from a database in which 30 images were used for training. The achieved retrieval rate is 73.93%. Iqbal and Aggarwal [6] had exploited the semantic interrelationships between different primitive image features by perceptual grouping to detect the presence of man-made structures. Their methodology retrieves building images based on these principles in a Bayesian framework. The system had a recall of maximum 80% and a precision of 83.72% for the class of images containing buildings. In content-based image retrieval system image representation is a challenging problem.

Attributed relational graph (ARG) [7] can be a powerful representation. Yu and Grimson [8] used ARG for image representation. It is a composition of vertices or attributed parts (color, shape, e.g.) and edges or attributed relations such as relative brightness, relative texture change, and relative positions. A subgraph of an ARG is called *configuration* which is very efficient for representing contextual information in an image. Their framework combined configurational and statistical approaches in image retrieval. Instead of representing an image by a set of *configurations* they came up with a vector-space structure or statistical feature-based representation deducted from the *configurations* making the concept of learning and prediction easier. Thus their method is enriched with the semantic description power of *configurations* and simple vector-space structure of statistical approaches.

SIMPLIcity (semantics sensitive-integrated matching for picture libraries) [9] is an efficient CBIR system, which uses semantic classification methods, wavelet-based approach for feature extraction, and integrated region matching based upon image segmentation. The system classifies images in categories like textured-nontextured and graph-photograph. This categorization enhances retrieval by permitting semantically adaptive searching methods and also narrowing down the search space. A similarity measure is developed using region matching scheme which integrates properties of all regions in an image. Experimentation results showed that SIMPLIcity is a faster, better, and robust method for CBIR. Some works [10–12] have been done for naturalness classification or man-made versus natural image classification. In this case, images are represented by their "spatial envelope" properties, including naturalness, openness, and roughness. However, robust indoor versus outdoor scene classification is a challenging problem in the sense that both kinds of images can have common man-made objects and content of images are more unconstrained. Luo and Boutell [10] tried to cope with this challenge by using over-complete independent component analysis (ICA) on the Fourier-transformed image to obtain sparse representation, serving for more accurate classification. Some approaches [11] used only texture orientation as a low-level feature to discriminate "city/suburb" images. In [12], it has been reported that high-level information can be inferred from low-level information and also high classification rate can be obtained from high-level feature set, whereas low-level feature gives low accuracy with low computational cost. A two-stage indoor/outdoor classification scheme has been attempted by Navid Serrano and Luo [12] using low-level features like texture and color. Images are divided into a number (powers of 2) of square blocks. Each of the blocks passes through color and texture feature extractor to be classified separately as indoor/outdoor blocks. And finally another classifier is used to classify the blocks into indoor or outdoor. The drawback of this method is that a fixed square blocking is applied to input images.

The method proposed in our paper segments the image using FCM based on features obtained using discrete wavelet transform to generate a set of segments which perceptually represents an indoor or outdoor image. We have used an unsupervised classifier (FCM) to segment the images such that it has no bias towards indoor or outdoor scenes. Unsupervised texture segmentation using FCM, based on features obtained from the two most commonly used multiresolution, multichannel filters: Gabor function and wavelet transform are described in [13]. A feature set has been derived from distinct regions and fed to a PNN (probabilistic neural network) for classification of the entire scene. The overall flowchart of the proposed method is given in Figure 1.

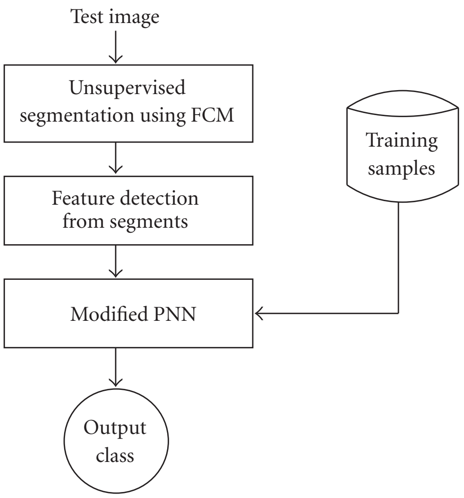

Test image
Unsupervised segmentation using FCM
Feature detection from segments
Modified PNN
Training samples
Output class

Figure 1: Block diagram of the proposed technique for scene classification.

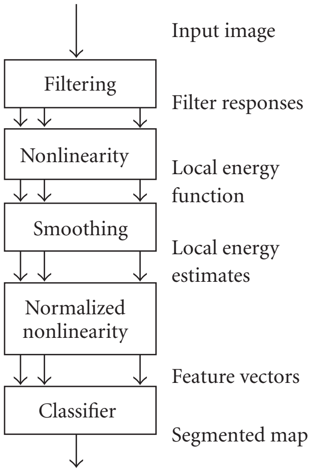


Figure 2: Stages of preprocessing for scene segmentation.

3. SCENE SEGMENTATION

In order to extract local features from the scene, the image is initially segmented using fuzzy C-means clustering [14] based on wavelet features [15]. We have used an unsupervised classifier (FCM) to segment the images such that it has no bias towards indoor or outdoor scenes. It is assumed that humans identify large parts of a scene for object recognition or scene understanding by analyzing a picture in modules [16]. Figure 2 shows the steps involved in image segmentation [13]. Each spectral band of the input image is filtered using discrete wavelet transform (Daubechies 8-tap and Haar filters). The absolute value of filter responses are smoothed by a Gaussian function. This is further normalized and the statistical features extracted for each spectral band (red, green, and blue) are concatenated to form an augmented feature vector which is used for clustering. The following subsections elaborate on the extraction of wavelet features, the postprocessing, and clustering using fuzzy C-means technique.

3.1. *Feature extraction using discrete wavelet transform (DWT)*

The discrete wavelet transform analyzes a signal based on its content in different frequency ranges. Therefore it is very useful in analyzing repetitive patterns such as texture [15, 17]. The 2D wavelet transform uses a family of wavelet functions and its associated scaling functions to decompose the original image into different subbands, namely, the low-low, low-high, high-low, and high-high (A, V, H, D, resp.) subbands. The decomposition process can be recursively applied to the approximation subband (A) to generate decomposition at the next level. Figures 3(a) and 3(b) show the level-2 dyadic decomposition of an image. The filter responses are postprocessed to compute the local energy estimates (as shown in Figure 4). The absolute value of a filter response $h_l^q(x,y)$ is convolved with a low-pass Gaussian post filter $g(x,y)$ to yield a post-filtered energy of the qth subband of lth filter as

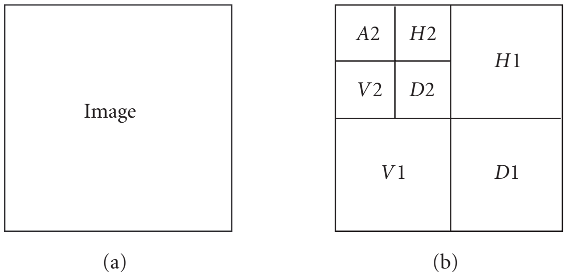


Figure 3: (a) Input image, (b) decomposition at level 2.

$$e_l^q(x,y) = \left|h_l^q(x,y)\right| * * g(x,y), \tag{1}$$

where

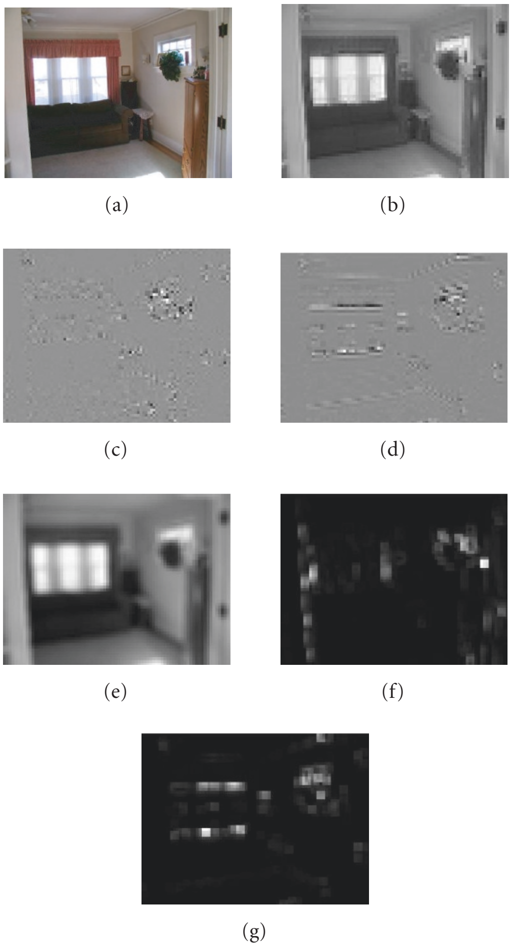
$$g(x,y) = \frac{1}{2\pi\sigma_2^2} e^{-[x^2+y^2]/2\pi\sigma_2^2}, \tag{2}$$

$**$ denotes 2D convolution, and $|\cdot|$ denotes absolute value. The feature vectors computed from the local window around a given pixel from the energy estimates are

(1) mean: $\mu = E[e_l^q(x,y)]$, of postprocessed A;
(2) variance: $\sigma = E[(e_l^q(x,y) - \mu)^2]$, of postprocessed V and H.

Here the $E[\cdot]$ is the expectation operator. The three wavelet components A, V, and H, for the green spectral band of the image shown in Figure 4(a), are shown in Figures 4(b)–4(d). The corresponding Gaussian postfiltered outputs are shown in Figures 4(e)–4(g). The final feature vector obtained for each pixel of an image can be expressed as

$$\begin{aligned}\mathbf{x}(x,y) = \Big[&\mu_{A_R}^d(x,y)\ \sigma_{V_R}^d(x,y)\ \sigma_{H_R}^d(x,y)\\ &\mu_{A_R}^h(x,y)\ \sigma_{V_R}^h(x,y)\ \sigma_{H_R}^h(x,y)\Big]^T,\end{aligned} \tag{3}$$



(a) (b) (c) (d) (e) (f) (g)

Figure 4: (a) Input image, (b)–(d) approximation, horizontal and vertical components, respectively, of the input image in (a), (e)–(g) energy map computed by postprocessing images in (b)–(d).

where $\mathbf{x}(x, y)$ is the feature vector, $\mu^{d}_{A_R}(x, y)$ is the estimated mean of the energy in the approximation subband obtained by filtering red spectral band of input image (using 8-tap Daubechies wavelet filter), and $\sigma^{h}_{V_R}(x, y)$ is variance of the estimated energy in the vertical subband (using Haar filter).

Similarly for each spectral band (red, green, and blue) mean of A and variance of V and H are computed for responses obtained using two wavelet filters (Daubechies, and Haar). Thus an eighteen-dimension feature vector is obtained by concatenating all features obtained using these combinations. Hence each pixel in the image is now represented by a feature in $\Re^{18}$. This is used to segment the image using an unsupervised method of segmentation, which is described in the following subsection.

3.2. Fuzzy c-means clustering

There are already a large number of supervised and unsupervised texture segmentation algorithms existing in the literature. The difference between supervised and unsupervised segmentation is that supervised segmentation assumes prior knowledge on the type of textures present in the image. We have used here the (unsupervised) fuzzy C-means clustering (FCM) algorithm [14] which is an iterative procedure. Given M input feature vectors $\mathbf{x}_m$, $m = 1, \dots, M$, the number of clusters C, where $2 \le C < M$, and the fuzzy weighting exponent z, $1 < z < \infty$, initialize the fuzzy membership function $u^{(0)}_{c,m}$ which is an entry of a $C \times M$ matrix $\mathbf{U}^{(0)}$. The following steps are iterated for increments of b.

(1) Calculate the fuzzy cluster centers $\mathbf{v}^{b}_{c}$ with

$$\mathbf{v}^{b}_{c} = \frac{\sum_{m=1}^{M} \left(u^{b}_{c,m}\right)^{z} \mathbf{x}_m}{\sum_{m=1}^{M} \left(u^{b}_{c,m}\right)^{z}}. \tag{4}$$

(2) Update $\mathbf{U}$ with

$$u^{b+1}_{c,m} = \left[\sum_{j=1}^{C} \left\langle \frac{\alpha_{c,m}}{\alpha_{j,m}} \right\rangle^{2/(z-1)}\right]^{-1}, \tag{5}$$

where $(\alpha_{j,m})^2 = \|\mathbf{x}_m - \mathbf{v}^{b}_{j}\|^2$ and $\|\cdot\|$ is any inner product-induced norm.

(3) Compare $\mathbf{U}^{b}$ with $\mathbf{U}^{(b+1)}$ in a convenient matrix norm. If $\|\mathbf{U}^{(b+1)} - \mathbf{U}^{(b)}\| \le \varepsilon$ ($\varepsilon = 10^{-5}$) stop, return to step 1.

The value of the weighting exponent z determines the fuzziness of the clustering decision. A smaller value of z, that is, z close to unity, will lead to a zero/one hard decision membership function, while a larger z corresponds to a fuzzier output. Figure 5(b) shows the segmented output for the image shown in Figure 5(a). Different shades of gray represent distinct clusters, where only the four significant (largest based on area) segments are considered. Figures 5(c)–5(f) show the bitmasks corresponding to the four major segments from the segmented image. In this work although the FCM-based clustering assigns disconnected image segments to the same cluster we consider disconnected segments of the same cluster as different segments. Regions near the boundary are not considered for further processing as they are often not completely available.

4. LOCAL FEATURE EXTRACTION

Local feature extracted from each of the major segment of the image are color, texture, and shape characteristics. Each type of feature is normalized and concatenated to form the augmented feature as

$$\mathbf{x} = \left[\mathbf{x}_{\text{color}} \;\; \mathbf{x}_{\text{texture}} \;\; \mathbf{x}_{\text{shape}}\right]^{T}. \tag{6}$$

In the following, each type of feature used for classification is discussed.

Color

For each segment of the image, the mean color values are taken as the feature

$$\mathbf{x}_{\text{color}} = \left[\mu_R \;\; \mu_G \;\; \mu_B\right]^{T}, \tag{7}$$

where μ is the mean for the red (R), green (G), and blue (B) bands.

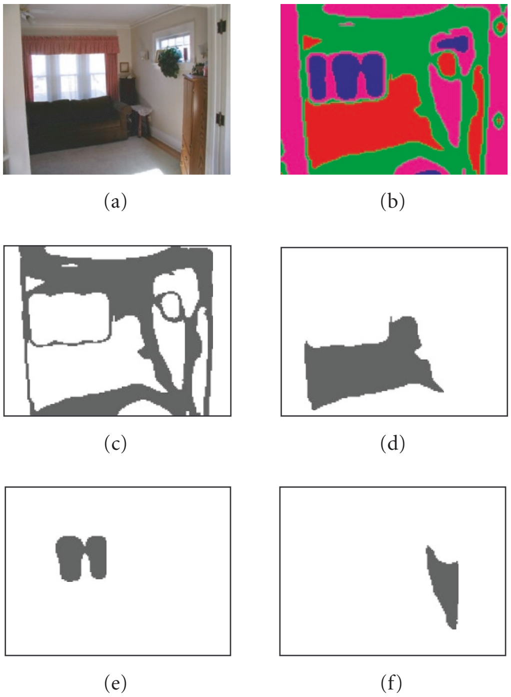

(a) (b) (c) (d) (e) (f)

Figure 5: (a) Input image, (b) the segmented output, (c)–(f) bitmasks for different connected image segments indicated by gray shade.

Texture

A feature vector for each segment is computed by taking mean of all the features associated with the pixels in a segment as

$$\tilde{\mu}^d_{A_R} = \frac{1}{P}\sum_{(x,y)\in\xi}\mu^d_{A_R}(x,y), \qquad \tilde{\sigma}^d_{A_R} = \frac{1}{P}\sum_{(x,y)\in\xi}\sigma^d_{V_R}(x,y),$$
$$\tilde{\sigma}^d_{A_R} = \frac{1}{P}\sum_{(x,y)\in\xi}\sigma^d_{H_R}(x,y), \tag{8}$$

where P is the cardinality of the set ξ of pixels in a segment s, of the image. Similarly, mean features are computed for other features mentioned in Section 3. The texture feature vector thus obtained is

$$\mathbf{x}_{\text{texture}} = \left[\tilde{\mu}^d_{A_R} \;\; \tilde{\sigma}^d_{V_R} \;\; \tilde{\sigma}^d_{H_R} \;\; \tilde{\mu}^h_{A_R} \;\; \tilde{\sigma}^h_{V_R} \;\; \tilde{\sigma}^h_{H_R} \;\; \ldots\right]^T. \tag{9}$$

Shape

Shape has been used as a feature for discriminating object classes. The Blobworld system [18] computes the area, eccentricity, and orientation of each region corresponding to an object. In this work, we use three shape features: eccentricity, compactness, and Euler number, to represent scene segments. The shape features are invariant to translation, rotation, and scaling. We consider such invariance important for obtaining a robust classification. Eccentricity and compactness are used as global parameters for MPEG-7 shape descriptors [19].

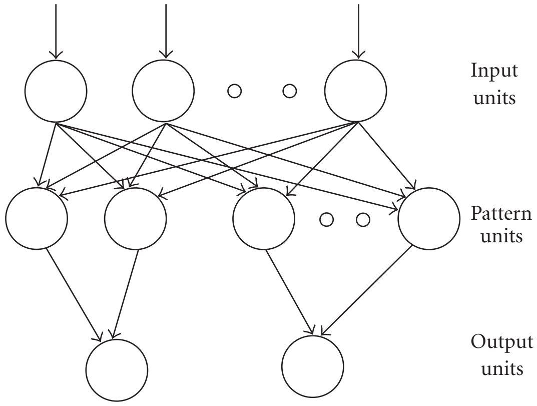


Figure 6: Probabilistic neural network architecture.

(1) Eccentricity is the ratio of the length of the longest chord of the shape to the longest chord perpendicular on it.

(2) Compactness is often defined as the ratio of squared perimeter and the area of an object:

$$x_{\text{compactness}} = \frac{(\text{Perimeter})^2}{\text{Area}}. \tag{10}$$

Compactness reaches the minimum in a circular object and approaches infinity in thin, complex objects.

(3) Euler number is used as the topological descriptor defined as the number of connected components minus the number of holes in the segmented regions.

The above-mentioned shape features are concatenated to form the shape feature vector

$$\mathbf{x}_{\text{shape}} = \left[x_{\text{eccentricity}} \;\; x_{\text{compactness}} \;\; x_{\text{Euler}}\right]^T. \tag{11}$$

5. CLASSIFICATION

5.1. Probabilistic neural network

The PNN model is based on Parzen's results on probability density function (PDF) estimators [20, 21]. PNN is a three-layer feedforward network consisting of input layer, a pattern layer, and a summation or output layer as shown in Figure 6. We wish to form a Parzen estimate based on K patterns each of which is n-dimensional, randomly sampled from c classes. The PNN for this case consists of n input units comprising the input layer, where each unit is connected to each of K pattern units; each pattern unit is, in turn, connected to one and only one of the c category units. The connection from the input to pattern units represents modifiable weights, which will be trained. Each category unit computes the sum of the pattern units connected to it. A radial basis function and a Gaussian activation function are used for the pattern nodes.

The PNN is trained in the following way. First, each pattern $\mathbf{x}$ of the training set is normalized to have unit length. The first normalized training pattern is placed on the input units. The modifiable weights linking the input units and the first pattern unit are set such that $\mathbf{w}_1 = \mathbf{x}_1$. Then, a single connection from the first pattern unit is made to the category unit corresponding to the known class of that pattern. The process is repeated with each of the remaining training patterns, setting the weights to the successive pattern units such that $\mathbf{w}_k = \mathbf{x}_k$ for $k = 1, 2, \ldots, K$. After such training we have a network which is fully connected between input and pattern units, and sparsely connected from pattern to category units. The trained network is then used for classification in the following way. A normalized test pattern $\mathbf{x}$ is placed at the input units. Each pattern unit computes the inner product to yield the net activation y,

$$y_k = \mathbf{w}_k^T \mathbf{x} \tag{12}$$

and emits a nonlinear function of y_k; each output unit sums the contributions from all pattern units connected to it. The activation function used is $\exp(\|\mathbf{x} - \mathbf{w}_k\|/\sigma^2)$. Assuming that both $\mathbf{x}$ and $\mathbf{w}_k$ are normalized to unit length, this is equivalent to using $\exp((y_k - 1)/\sigma^2)$. As the number of segments obtained differs from one scene to another, the feature-set representation of the scene is of varying dimension. Therefore a modified PNN is used for classifying the variable dimension feature sets.

5.2. *Modified PNN*

In our work, the second layer (i.e., pattern layer) must have

$$\widetilde{K} = \sum_{i=1}^{I} S_i \tag{13}$$

units, where I is the total number of training images for both indoor and outdoor classes and S_i denotes number of segments in ith image. Here we consider different segments in training scenes to train our network. To classify a test scene, each segment of the test image is compared with each unit in the pattern layer. The distance between feature vector associated with the segment(s) of the test image and the weight vector associated to the pattern unit is computed as

$$d_k = \min \|\mathbf{x}(s) - \mathbf{w}_k\|, \tag{14}$$

where d_k is the distance between the closest segment (sth segment) of the test image to the kth weight vector. We find the closest segment of the test image to each one of the training segments. The activation function used here is $\exp(d_k/\sigma^2)$. The value of σ is found to be 0.07 by trial and error method. The output layer contains two units, one of them connects to all the units in the pattern layer containing segments corresponding to indoor scene and the other connects to all remaining units in pattern layer (units corresponding to outdoor scenes). For an unknown test scene, each output unit sums the contributions from all pattern units connected to it. The output unit with the highest value wins. In case of a competition between the two output units, the one with the most number of closely associated segments (based on d_k in (14)) will be considered for obtaining a crisp classifier decision.

Table 1: Indoor versus outdoor classification accuracy (%) on IITM-SCID2 and Benchmark-2.

Feature set	IITM SCID2		Benchmark-2	
	Indoor	Outdoor	Indoor	Outdoor
Shape	63.5	66.5	26.1	89.4
Color	94.0	53.5	79.5	90.7
Texture	94.0	86.9	90.1	82.6
Shape + color	89.2	71.5	75.2	95.7
Shape + texture	90.4	83.8	83.9	89.4
Color + texture	**94.0**	**90.8**	**89.4**	**85.1**
Shape + color + texture	89.6	83.1	84.5	89.4

Table 2: Comparison of various methods for indoor versus outdoor classification accuracy (%).

Methods	IITM SCID2		Benchmark-2	
	Indoor	Outdoor	Indoor	Outdoor
Proposed (color + texture)	94.0	90.8	89.4	85.1
Edge straightness (rule-based)	71.0	72.5	85.0	80.0
Edge straightness (k-NN)	65.5	66.5	78.9	87.9

6. EXPERIMENTAL RESULTS

The proposed scene classification method is tested on the IITM-SCID2 (scene classification image database) [22] and part of the image database provided by the authors in [2] (we call this Benchmark-2). The IITM-SCID2 database consists of 902 indoor and outdoor images together, out of which 193 indoor and 200 outdoor images are used for training, and 249 indoor and 260 outdoor images are used for testing. The Benchmark-2 database consists of around 522 indoor and outdoor images together, out of which 100 images per class were used for training and 161 images per class were used for testing. The features extracted were normalized across the entire training and testing sets and concatenated to form the augmented feature vector for each combination. This augmented feature vector is used during training and testing the modified PNN.

Table 1 shows the classification performance of the proposed method with different combinations of color, texture, and shape features on the IITM-SCID2 and Benchmark-2 databases. It can be observed that the combination of color and texture features perform better than all other combinations of features put together for both databases. It can also be noted that out of the three different types of features used individually, the textural features perform significantly better than shape- and color-based features for both databases. The performance of shape features is particularly good for outdoor scenes in Benchmark-2, but the performance of color features do not follow a trend for both databases due to the differences in color variations in both the databases. For indoor scenes in case of Benchmark-2 the shape features provide poor results. This leaves the scope of exploring better shape measures for classification.

Table 2 compares the classification performance of the proposed method and our implementation of the methods proposed in [2] on IITM-SCID2 and Benchmark-2. It can be observed that the proposed method performs significantly

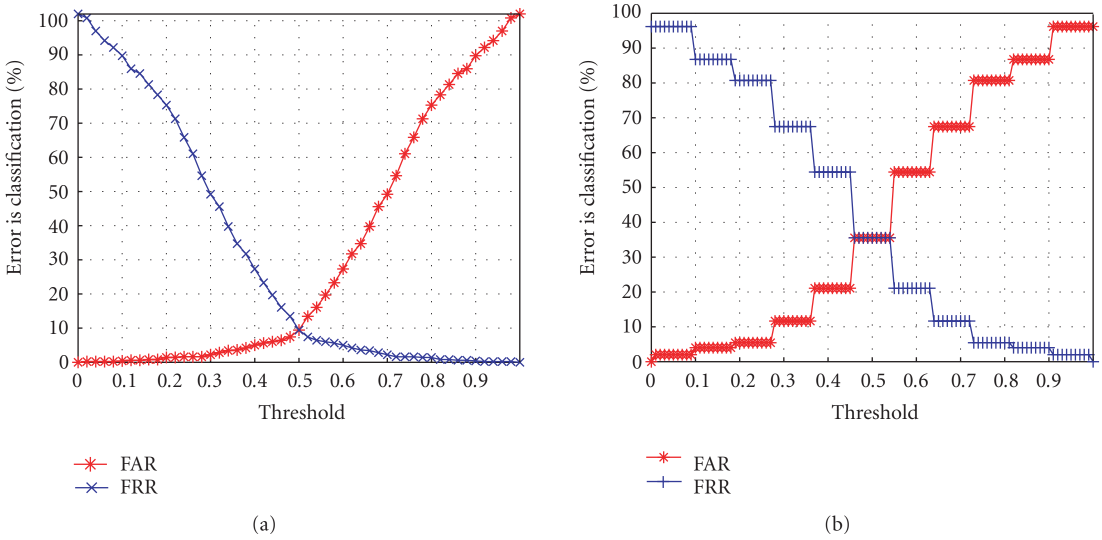
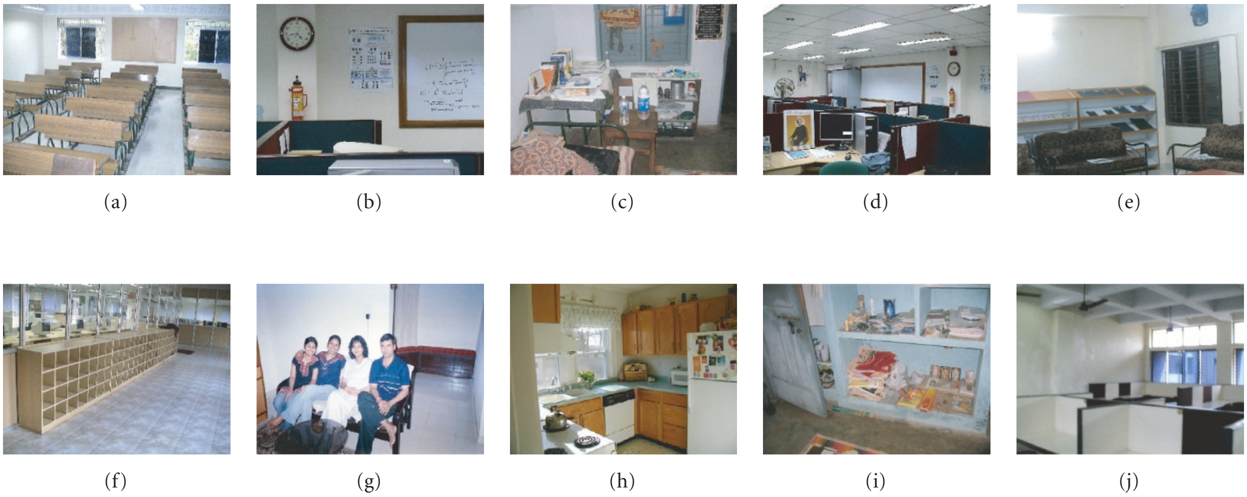


FIGURE 7: False acceptance and false rejection rates (FAR and FRR): (a) for the proposed method, and (b) k-NN method, on IITM-SCID2.


(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

FIGURE 8: Examples of correctly classified indoor images (from IITM-SCID2).

better than both the methods proposed in [2] on IITM-SCID2 and Benchmark-2. We have obtained 83% overall classification accuracy on Benchmark-2 using our implementation of the method proposed in [2], which is near to that (87%) quoted in [2].

Figure 7 shows the FAR and FRR values for the proposed method and the method proposed in [2]. It can be observed that the equal error rate (EER) for the proposed method is 9.4%. This is significantly lesser than EER obtained for our implementation of [2] which is 35.5%. Figures 8 and 9 show some of the correctly classified indoor and outdoor scenes, respectively, from IITM-SCID2. Figure 10 shows the indoor images that were incorrectly classified as outdoor class from IITM-SCID2. This may be due to the inadequacy of the training images to provide the variability necessary to correctly classify the segments of the test image. Figure 11 shows the outdoor images that were incorrectly classified as indoor scenes from IITM-SCID2. It can be observed that most of these images have characteristics similar to indoor images such as flat-shaded walls with smooth textures and image segments with straight borders. Figures 12 and 13 show some of the correctly classified indoor and outdoor scenes, respectively, from Benchmark-2. Figures 14 and 15 show some of the incorrectly classified indoor and outdoor scenes, respectively, from Benchmark-2.

7. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a method for indoor versus outdoor scene classification. We have represented the image

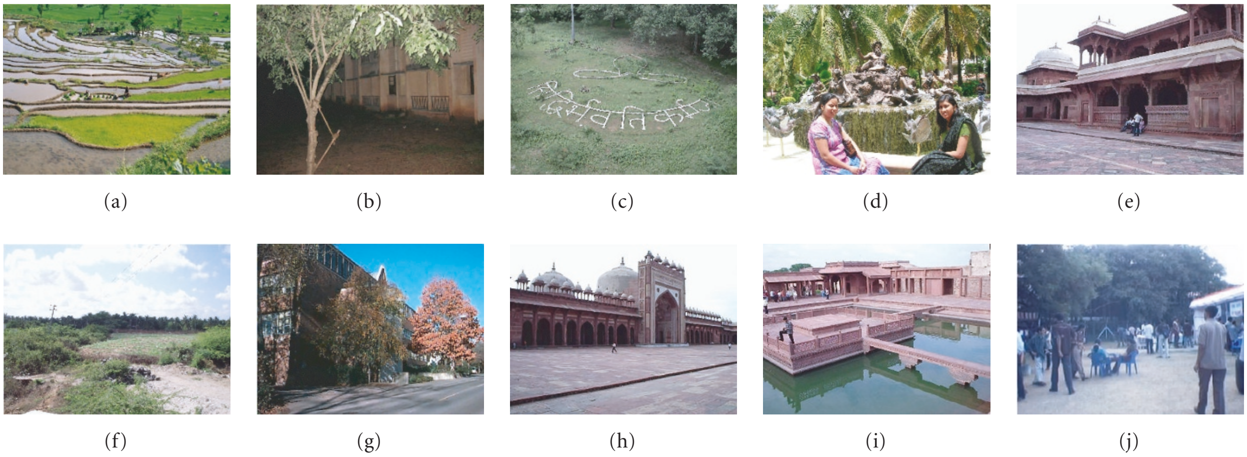

(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)

Figure 9: Examples of correctly classified outdoor images (from IITM-SCID2).

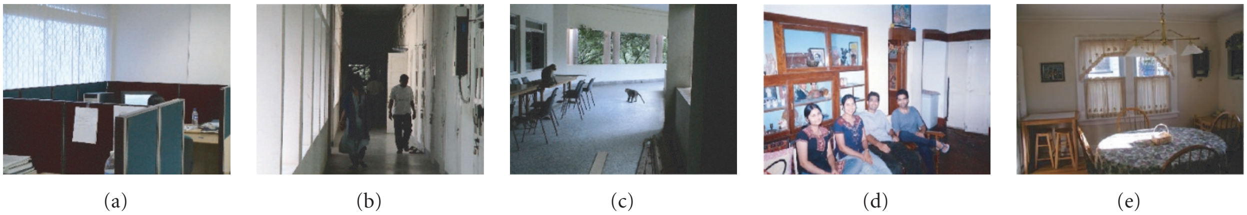

(a) (b) (c) (d) (e)

Figure 10: Examples of indoor images misclassified as outdoor scenes (from IITM-SCID2).

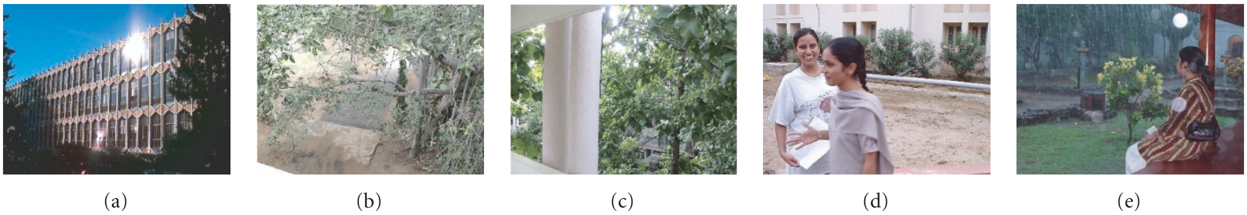

(a) (b) (c) (d) (e)

Figure 11: Examples of outdoor images misclassified as indoor images (from IITM-SCID2).

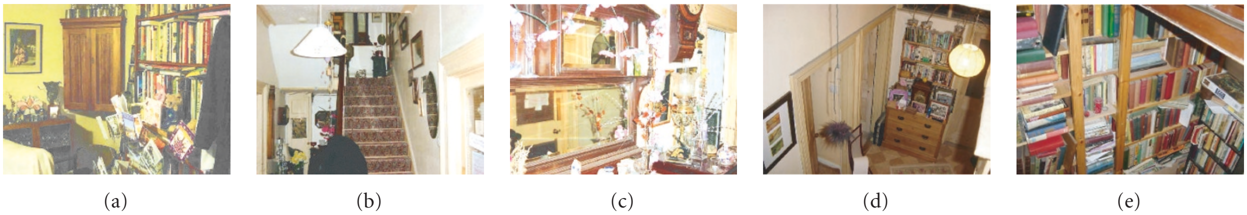

(a) (b) (c) (d) (e)

Figure 12: Examples of correctly classified indoor images (from Benchmark-2).

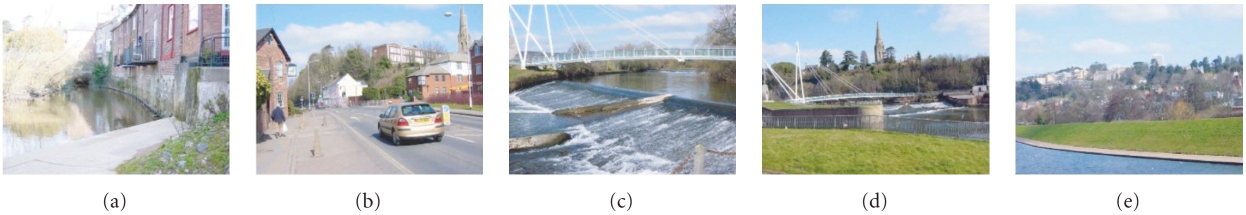

(a) (b) (c) (d) (e)

Figure 13: Examples of correctly classified outdoor images (from Benchmark-2).

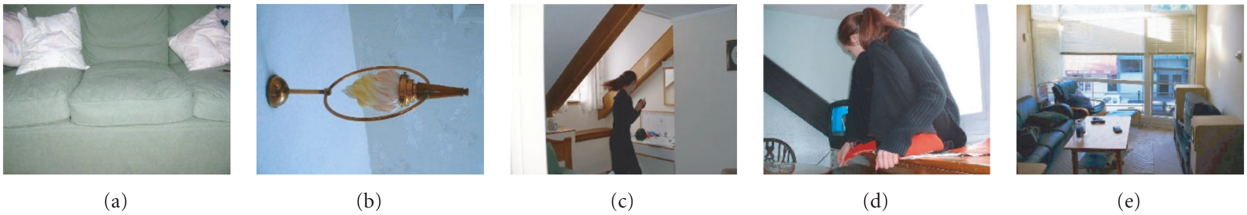


Figure 14: Examples of indoor images misclassified as outdoor scenes (from Benchmark-2).

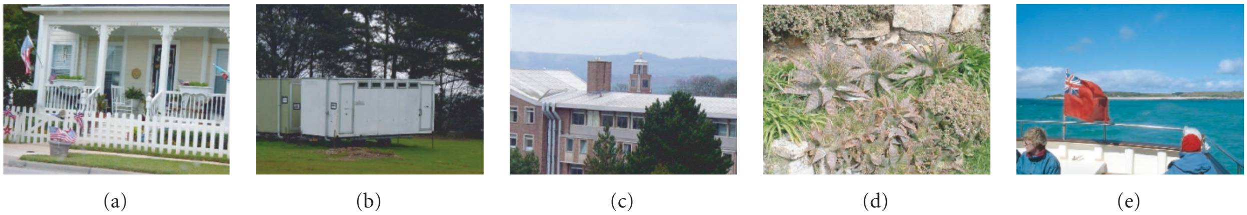


Figure 15: Examples of outdoor images misclassified as indoor images (from Benchmark-2).

using a feature set with varying number of feature vectors each describing the local color, shape, and textural properties of the image segments. In order to classify a variable dimension feature set, a modified PNN is used to overcome the problem of varying number of feature vectors, of the feature set, corresponding to the number of segments in the scene. We have tested the proposed scene classification technique on the IITM-SCID2 database and observed that the textural features based on the DWT subbands dominates other features such as shape and color. Future work includes exploring the use of a richer feature set based on other properties such as moments, edge ratio, and straightness of the edge. The modified PNN used in this work can be further extended to scene matching for image-querying applications.

REFERENCES

[1] E. Saber and A. M. Tekalp, “Integration of color, edge, shape, and texture features for automatic region-based image annotation and retrieval,” *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 684–700, 1998.

[2] A. Payne and S. Singh, “Indoor vs. outdoor scene classification in digital photographs,” *Pattern Recognition*, vol. 38, no. 10, pp. 1533–1545, 2005.

[3] A. K. Jain and A. Vailaya, “Image retrieval using color and shape,” *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, 1996.

[4] A. Vailaya, A. Jain, and H. J. Zhang, “On image classification: city images vs. landscapes,” *Pattern Recognition*, vol. 31, no. 12, pp. 1921–1935, 1998.

[5] Q. Iqbal and J. K. Aggarwal, “Image retrieval via isotropic and anisotropic mappings,” in *Proceedings of IAPR Workshop on Pattern Recognition in Information Systems*, pp. 34–49, Setubal, Portugal, July 2001.

[6] Q. Iqbal and J. K. Aggarwal, “Applying perceptual grouping to content-based image retrieval: building images,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’99)*, vol. 1, pp. 42–48, Fort Collins, Colo, USA, June 1999.

[7] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, Reading, Mass, USA, 1992.

[8] H. Yu and W. E. L. Grimson, “Combining configurational and statistical approaches in image retrieval,” in *Proceedings of the 2nd IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, vol. 2195 of *Lecture Notes in Computer Science*, pp. 293–300, Beijing, China, October 2001.

[9] J. Z. Wang, J. Li, and G. Wiederhold, “Simplicity: semantics-sensitive integrated matching for picture libraries,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

[10] J. Luo and M. Boutell, “Natural scene classification using overcomplete ICA,” *Pattern Recognition*, vol. 38, no. 10, pp. 1507–1519, 2005.

[11] M. M. Gorkani and R. W. Picard, “Texture orientation for sorting photos “at a glance”,” in *Proceedings of the 12th International Conference on Pattern Recognition (ICPR ’94)*, vol. 1, pp. 459–464, Jerusalem, Israel, October 1994.

[12] A. S. Navid Serrano and J. Luo, “A computationally efficient approach to indoor/outdoor scene classification,” in *Proceedings of the International Conference on Pattern Recognition (ICPR ’02)*, vol. 4, pp. 146–149, Quebec City, Quebec, Canada, August 2002.

[13] S. G. Rao, M. Puri, and S. Das, “Unsupervised segmentation of texture images using a combination of gabor and wavelet features,” in *Proceedings of the 4th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP ’04)*, pp. 370–375, Kolkata, India, December 2004.

[14] M. F. A. Fauzi and P. H. Lewis, “A fully unsupervised texture segmentation algorithm,” in *Proceedings of the British Machine Vision Conference (BMVC ’03)*, pp. 519–528, Norwich, UK, September 2003.

[15] E. Salari and Z. Ling, “Texture segmentation using hierarchical wavelet decomposition,” *Pattern Recognition*, vol. 28, pp. 1819–1824, 1995.

[16] I. E. Gordon, *Theories of Visual Perception*, Psychology Press, New York, NY, USA, 3rd edition, 2004.

[17] C.-S. Lu, P.-C. Chung, and C.-F. Chen, "Unsupervised texture segmentation via wavelet transform," *Pattern Recognition*, vol. 30, no. 5, pp. 729–742, 1997.

[18] C. Carson, M. Thomas, M. Belongie, J. Hellerstein, and J. Malik, "Blobworld: a system for region based image indexing and retrieval," in *Proceedings of the 3rd International Conference on Visual Information Systems*, Amsterdam, The Netherlands, June 1999.

[19] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standarization*, Kluwer Academic, Boston, Mass, USA, 2003.

[20] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[21] P. E. H. Richard, O. Duda, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2004.

[22] "IIT Madras Scene Classification Image Database (SCID)," http://vplab.cs.iitm.ernet.in/SCID/.

Lalit Gupta is pursuing his M.S. degree at the Department of Computer Science and Engineering, Indian Institute of Technology Madras. Currently he is working on image-texture analysis. His research interests include computer vision and pattern recognition. He has published one paper in national conference.

Vinod Pathangay received the M.S. degree from Indian Institute of Technology Madras in 2004 and currently pursuing the Ph.D. degree there with fellowship from Infosys Foundation. His current research interests are computer vision and pattern recognition. He has published one paper in national conference.

Arpita Patra is pursuing her M.S. degree at the Department of Computer Science and Engineering, Indian Institute of Technology Madras under the guidance of Dr. Sukhendu Das. Currently she is working on face recognition and multimodal biometry. During her M.S. degree she has completed a project named "Multimodal biometric-based secured access system using face and fingerprint recognition." Her research interests include computer vision, image processing, and statistical pattern recognition.

A. Dyana received the M.Tech. degree from Manonmanium Sundaranar University, Tirnelveli, India in information technology and currently is pursuing the Ph.D. degree from Indian Institute of Technology Madras. Her research interests include computer vision and image compression. She has published one paper in national conference.

Sukhendu Das is currently working as an Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, India. He completed his B.Tech. degree from Indian Institute of Technology Kharagpur from the Department of Electrical Engineering in 1985 and M.Tech. degree in the area of computer technology from Indian Institute of Technology Delhi in 1987. He then obtained his Ph.D. degree from Indian Institute of Technology Kharagpur in 1993. His current areas of research interests are visual perception, computer vision, digital image processing and pattern recognition, computer graphics, artificial neural networks, and computational science and engineering. He has been in the faculty of the Department of Computer Science and Engineering, Indian Institute of Technology Madras, India since 1989. He has also worked as a Visiting Scientist in the University of Applied Sciences, Pforzheim, Germany, for postdoctoral research work, from December 2001 till May 2003. He has guided one (currently guiding four) Ph.D. student and several M.S. (currently guiding eight), M.Tech, and B. Tech students. He had completed several international and national sponsored projects and consultancies, both as principle and coinvestigators. He has published more than 50 technical papers in international and national journals and conferences. He has received one best paper and a best design contest award.