Defence Science Journal, Vol. 68, No. 2, March 2018, pp. 129-137, DOI : 10.14429/dsj.68.10504

Indoor Scene Recognition for Micro Aerial Vehicles Navigation using Enhanced-GIST Descriptors

B. Anbarasu* and G. Anitha

Madras Institute of Technology Campus, Anna University, Chennai - 600 044, India
**E-mail*: *avianbu@gmail.com*

ABSTRACT

An indoor scene recognition algorithm combining histogram of horizontal and vertical directional morphological gradient features and GIST features is proposed in this paper. New visual descriptor is called enhanced-GIST. Three different classifiers, k-nearest neighbour classifier, Naïve Bayes classifier and support vector machine, are employed for the classification of indoor scenes into corridor, staircase or room. The evaluation was performed on two indoor scene datasets. The scene recognition algorithm consists of training phase and a testing phase. In the training phase, GIST, CENTRIST, LBP, HODMG and enhanced-GIST feature vectors are extracted for all the training images in the datasets and classifiers are trained for these image feature vectors and image labels (corridor-1, staircase-2 and room-3). In the test phase, GIST, CENTRIST, LBP, HODMG and enhanced-GIST feature vectors are extracted for each unknown test image sample and classification is performed using a trained scene recognition model. The experimental results show that indoor scene recognition algorithm employing SVM with enhanced GIST descriptors produces very high recognition rates of 97.22 per cent and 99.33 per cent for dataset-1 and dataset-2, compared to kNN and Naïve Bayes classifiers. In addition to its accuracy and robustness, the algorithm is suitable for real-time operations.

Keywords: Micro aerial vehicle; Indoor navigation; Scene recognition; Support vector machines; Classification algorithms; Video signal processing

1. INTRODUCTION

Indoor scene recognition[1] is one of the most fundamental task in computer vision and robotics. Indoor navigation of micro aerial vehicles (MAV) is a critical issue because GPS is not available in an indoor environment. For MAV to fly autonomously in indoor environment, it has to identify the type of the environment and based on that it has to follow a suitable navigation strategy. MAV is a rotary wing autonomous or remotely piloted aircraft. The MAV used in this work is a Parrot AR drone and is shown in Fig. 1. Quattoni and Torralba[1] proposed an indoor scene recognition algorithm by combining regions of interest (ROI) and global GIST features. Oliva[2], *et al.* proposed a computational scene recognition model for recognizing real world scenes based on the estimation of spatial envelope that represents the global structure of a scene. GIST was proposed[3] as a descriptor based on building the gist of the scene from global features. Wu and Rehg[4] proposed CENTRIST descriptor for indoor scene categorization. Ojala[5], *et al.* presented a computationally efficient multi-resolution approach based on 'uniform' local binary patterns to deal with gray-scale and rotation invariant texture classification. Vailaya[6], *et al.* proposed the local features for a specific high-level classification problem (city images vs. landscapes). Recently, object detection is used as an intermediate semantic representation for indoor scene recognition[7]. A metric function that explores the spatial distribution of indoor scenes are explained[8].

Fornoni[9], *et al.* combined saliency-driven perceptual pooling with a simple spatial pooling scheme to recognise indoor scenes. A weighted hypergraph learning based indoor scene classification is presented[10]. Khan[11], *et al.* proposed mid-level convolutional features for indoor image classification. Anbarasu[12], *et al.* has proposed a frontal obstacle detection and collision avoidance for MAV using ultrasonic sensors. Individual descriptor such as GIST, CENTRIST, HODMG and LBP does not encode both the spatial envelope and enhanced

Figure 1. Parrot AR drone2 quadrotor.

Received : 06 September 2016, Revised : 16 December 2017
Accepted : 01 January 2018, Online published : 13 March 2018

boundary information in indoor scenes. To overcome this shortcoming, new visual descriptor called enhanced-GIST descriptor is proposed for indoor scene recognition.

2. THE PROPOSED FRAMEWORK FOR INDOOR SCENE RECOGNITION

Proposed method as shown in Fig. 2 is designed to process the videos acquired by MAV and make the MAV to recognise the indoor environment such as corridor, staircase and room. The scene recognition model contains two stages i.e., training stage and testing stage. In the first step, for each training image the global image features and low-level features are extracted. Then combine both GIST and histogram of directional morphological gradients (HODMG) features in the next step. In the end of the training stage, classifiers are learned using these image feature vectors and image labels. In the Testing Stage, for each frame acquired by the MAV, enhanced GIST features are extracted and fed to the classifiers for indoor scene classification.

Using the trained model, the classifier will classify the type of indoor scenes. Our main focus is to classify the inter-class of 3 types of indoor scenes like corridor, staircase and room for the navigation of the MAV inside the building because the navigation rules for these three scenes would be different and would encompass majority of indoor navigation.

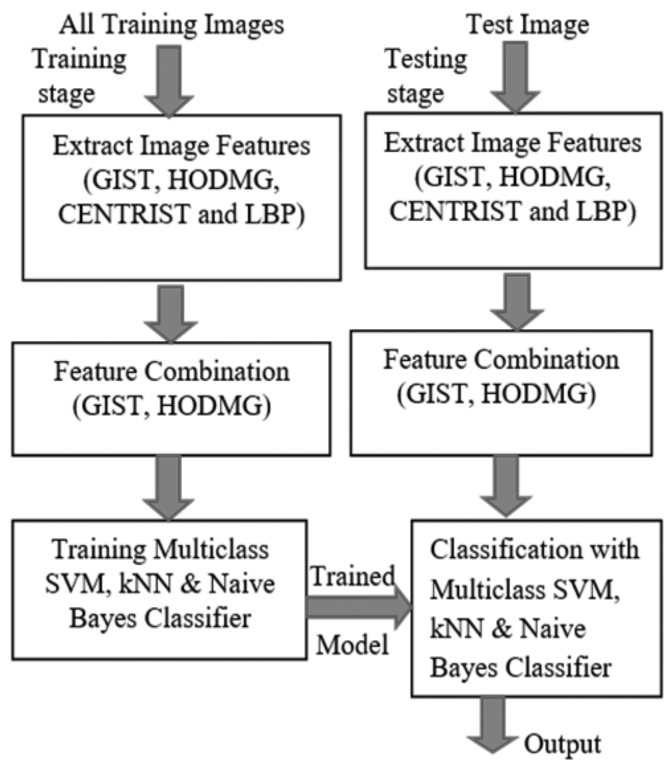


Figure 2. The block diagram of the proposed method.

3. IMAGE FEATURE EXTRACTION

3.1 GIST

GIST scene descriptor was proposed by Oliva[3], *et al.* In the image pre-processing stage, the RGB image with a resolution of 1280 × 720 pixels are converted into a grayscale image with a resolution of 256 × 256 pixels. Next, grayscale image was filtered by 32 Gabor filters at 4 scales and 8 orientations, to produce 32 feature maps. The 32 feature maps obtained are divided into 16 regions (4 × 4 grids) and finally the Gabor filtered outputs are averaged within each region to produce a 512 (16 × 32) dimensional GIST descriptor. The GIST descriptors extracted for the input frames of video acquired from the MAV is shown in Fig. 3.

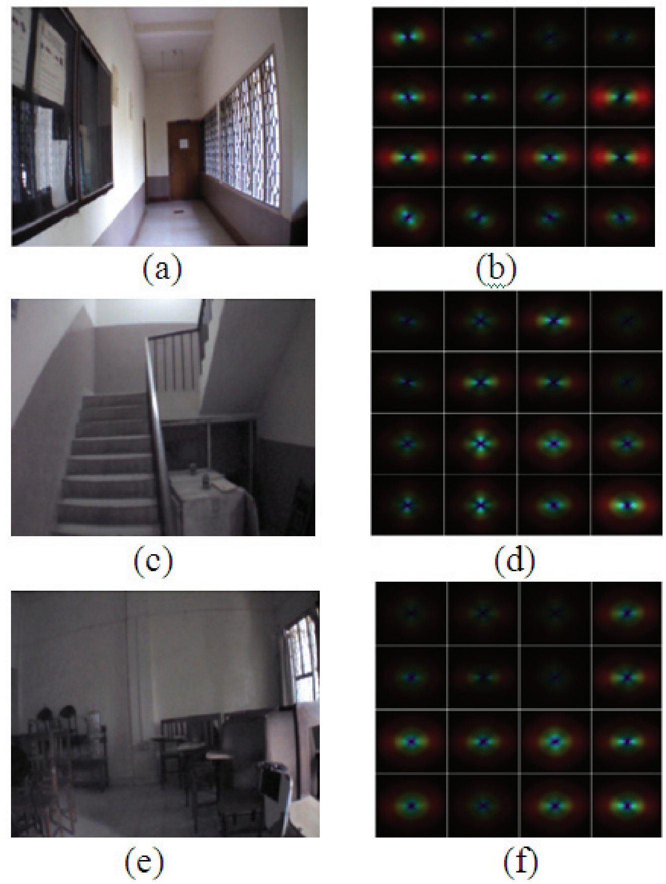


Figure 3. Input frames and their GIST descriptors for three indoor scenes: Corridor ((a) & (b)), Staircase ((c) & (d)), and Room ((e) & (f)).

3.2 Histogram of Directional Morphological Gradient

Directional morphological gradients[13,14] are computed for horizontal and vertical directions by using line segments as structuring element that is symmetric with respect to the neighbourhood center as shown in Figure 4. Directional gradients $gL_\alpha(f)$ for a given direction α can be computed as follows:

$$gL_\alpha(f) = \delta L_\alpha(f) - \varepsilon L_\alpha(f) \quad (1)$$

where $\delta L_\alpha(f)$ is the dilated image; $\varepsilon L_\alpha(f)$ is the eroded image and L is the structuring element. Horizontal and vertical directional gradients of the three indoor scenes (corridor, staircase, and room) are extracted from the input videos as illustrated in Fig. 4(a) - 4(f). Extracted HODMG is a 512-dimensional descriptor. Extracted GIST features are combined with HODMG features to produce a 1024-dimensional enhanced GIST feature descriptor.

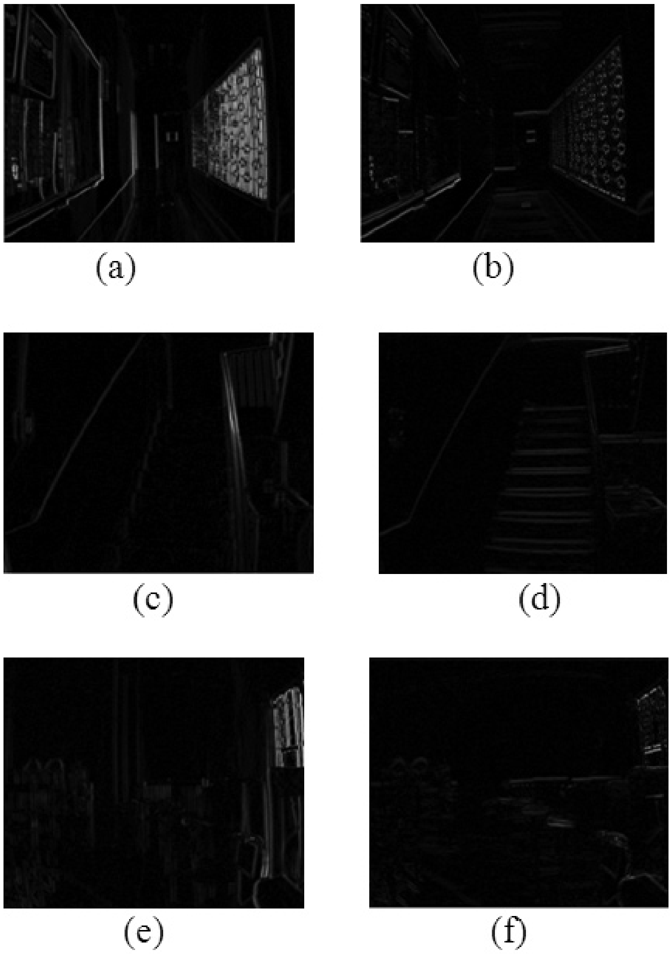


Figure 4. Horizontal and vertical directional gradients of indoor scenes: : Corridor ((a) & (b)), Staircase ((c) & (d)), and Room ((e) & (f)).

3.3 Centrist

Centrist is a new visual descriptor[4] that encodes the structural properties within an image and suppresses detailed textural information. Grayscale indoor image with resolution of 256 × 256 pixels are converted into Census Transformed image by comparing the intensity value of centre pixel to the intensity values of pixels in a 3 × 3 neighbourhood, and the obtained census transformed (CT) value as shown in Fig. 5.

If the neighbouring pixel value is greater than the center pixel value, bit '0' is assigned to the neighbouring pixel; otherwise the pixel value is set to '1'. Finally, the histogram of CT values obtained are used as visual descriptor. The CT images are as shown in Fig. 6. The CENTRIST (Not using PCA) is a 256-dimensional feature descriptor.

3.4 Local Binary Pattern

Local binary pattern[5] is obtained by assigning a binary number based on thresholding the 3 × 3 pixel neighbourhood with the pixel value at the centre as shown in Fig 7. If the centre pixel value is greater than or equal to the neighbouring pixel (3 × 3 pixel neighbourhood), binary value of '1' is assigned to the neighbouring pixel; otherwise, it is assigned a binary value of '0'.

66	68	68
66	67	67
67	66	67

⇨

1	0	0
1		1
1	1	1

⇨ $(10011111)_2$ =CT=159

Figure 5. Census transformed value.

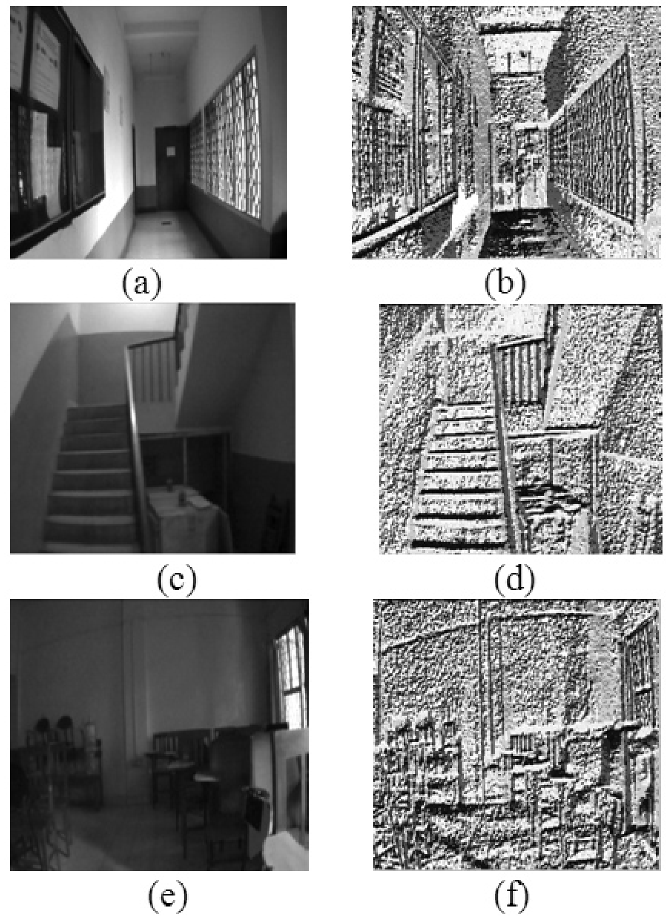


Figure 6. Input image and their census transformed image for indoor scenes: Corridor ((a) & (b)), Staircase ((c) & (d)), and Room ((e) & (f)).

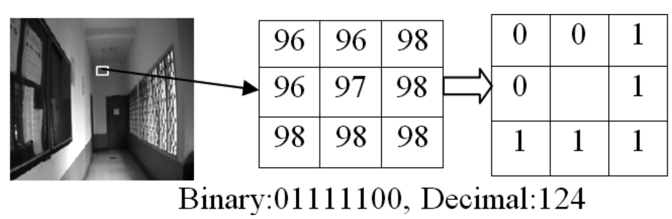


Figure 7. Illustration of basic LBP operator.

Finally, the binary digits are collected from the neighbouring pixels and converted to its decimal equivalent. The $LBP_{P,R}$ operator for a circular neighbourhood can be computed as follows:

$$LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{P-1} s(g_n - g_c)2^n \quad (2)$$

where x_c, y_c denote the locations of the central pixel, n denotes the intensity value of the nth neighbouring pixels, R

is the radius, and $s(x)$ is the unit step function, which returns $s(x)=1$ if $x \geq 0$, otherwise $s(x)=0$.

The rotation invariant uniform LBP value can be computed as follows:

$$LBP_{P,R}^{riu2}(x_c, y_c) = \begin{cases} \sum_{n=0}^{P-1} s(g_n - g_c), U(LBP_{P,R}) \leq 2 \\ P+1, otherwise \end{cases} \quad (3)$$

where subscript P denotes the circular neighbourhood of sampling points, subscript R denotes the radius of the circle and superscript *riu2* stands for using only rotation invariant uniform patterns. The histogram of local binary pattern feature is computed over the circular neighbourhood.

3.5 Indoor Datasets

The Dataset-1 contains a total of 252 images of 3 indoor scene classes chosen from MIT-67 indoor scene dataset. The MIT-67 (benchmark dataset) indoor scene classification dataset (15620 images) contains 67 indoor scene categories like airport inside, art studio, corridor, staircase, living room, bedroom, locker room, hospital room, etc. Specifically, 84 images per class (corridor and staircase) are considered; in dataset-1, 60 images and 24 images are used for training and testing the classifier, respectively. For room category 5 images per class and 2 images per class (bathroom, bedroom, children room, class room, computer room, dining room, game room, hospital room, living room, locker room, meeting room and operating room) are considered for training and testing respectively. For room category, 60 images and 24 images are used for training and testing, respectively. We have chosen dataset-1 (images from MIT-67 dataset) to test our algorithm on 3 image categories (corridor, staircase and room) mainly because it contains challenging and confusing images in the indoor image categories. Dataset-2 have 450 images of 3 indoor scene classes. Specifically, 150 images per class (corridor, staircase and room) are considered; in dataset-2, 100 images and 50 images are used for training and testing the classifier, respectively. The 100 training images per class are collected using the internet. For testing images out of 50 images per class, the first 20 images are collected using the internet and the remaining 30 images are the frames extracted from the transmitted video from the MAV.

Hence in dataset-2, 300 images and 150 images are used for training and testing the classifier, respectively. We have chosen dataset-2 to test our algorithm on image frames acquired from the MAV to validate the performance of the proposed algorithm for real time operations. Sample images from dataset-1 and dataset-2 are shown in Fig. 8.

4. CLASSIFIER

Three different classifiers SVM classifier, k-NN classifier and Naïve Bayes classifier are employed for the classification of indoor scenes.

4.1 SVM CLASSIFIER

One-Against-All (OAA) SVM[15,16] method is used in this study. In OAA method, k SVM models will be constructed to compare each indoor class with all other indoor classes of scenes for a k indoor classes. A training dataset $\{x_i, y_i\}_{i=1}^{l}$, where $x_i \in R^n, i=1,......,l$ and $y_i \in \{1,......,k\}$ represent the class of x_i, then the m^{th} SVM parameters can be determined by solving the following equation:

$$\min_{w^m, b^m, \xi^m} \quad \frac{1}{2}(w^m)^T w^m + C\sum_{i=1}^{l} \xi_i^m$$

$$\begin{aligned} &(w^m)^T \phi(x_i) + b^m \geq 1 - \xi_i^m, if \quad y_i = m, \\ &(w^m)^T \phi(x_i) + b^m \leq -1 + \xi_i^m, if \quad y_i \neq m, \\ &\xi_i^m \geq 0, i = 1,.....,l \end{aligned} \quad (4)$$

where ξ_i and C denotes the positive slack variable and regularization parameter (penalty value), respectively. By solving Eqn. (4), k decision functions are obtained as follows: $(w^1)^T \phi(x) + b^1, \ldots. (w^k)^T \phi(x) + b^k$.

Finally indoor scene class x can be determined based on the

Figure 8. Sample images for 3 scene categories from dataset-1 and dataset-2: corridor, staircase and room (from top to bottom).

maximum score of the decision function as follows

$$x \equiv \arg\max_{m=1,....,k}((w^m)^T \phi(x) + b^m) \quad (5)$$

4.2 kNN Classifier

For the kNN classifier[17,18], the Euclidean distance was selected as the distance metric expressed as follows:

$$d(r,a) = \sqrt{(r_1 - a_1)^2 + (r_2 - a_2)^2 + ... + (r_n - a_n)^2} \quad (6)$$

where r ($(r_1, r_2, ...r_n)$ and a $(a_1, a_2, a_3.......a_n)$ are n-dimensional vectors.

In the kNN classifier training phase, store the extracted feature vectors of the training samples and the corresponding class labels. In the classification phase the feature vectors of the testing set is classified based on the value of k (user defined constant) between 1 and 25 and the distance function.

4.3 Naive Bayes Classifier

In Naïve Bayes classifier[19] given a training set, and to classify a new instance expressed by the feature vector values $(a_1, a_2, a_3.......a_n)$, then the NBC will classify the target class by assigning the most probable target value (RMAP) for the new test sample (instance) expressed as

$$R_{MAP} = \arg\max_{R_j \in R} P(R_j \mid a_1, a_2, ...a_n) \quad (7)$$

Using Bayes theorem Eqn. (7) can be rewritten as

$$R_{MAP} = \arg\max_{R_j \in R} \frac{P(a_1, a_2, ...a_n \mid R_j)P(R_j)}{P(a_1, a_2, ...a_n)} \quad (8)$$

$$R_{MAP} = \arg\max_{R_j \in R} P(a_1, a_2, ...a_n \mid R_j)P(R_j) \quad (9)$$

Now the two terms in Eqn. (9) are estimated based on the training data. The probability of attributes $(a_1, a_2, a_3.......a_n)$ is the product of the probability of each feature:

$$P(a_1, a_2, ...a_n \mid R_j) = \prod_{i=1}^{n} P(a_i \mid R_j) \quad (10)$$

The univariate Kernel Density Estimation[20] can be expressed as

$$P(a) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{a - a_i}{h}\right) \quad (11)$$

where K(.), a and a_i are the density kernel, test instance point and the training instance point respectively. The univariate Gaussian kernel can be expressed as

$$K(\xi) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{\xi^2}{2}} \quad (12)$$

Thus, Eqn. (9) could be rewritten as follow to define the Naïve Bayes classification approach.

$$R_{NB} = \arg\max_{R_j \in R} P(R_j) \prod_{i=1}^{n} P(a_i \mid R_j) \quad (13)$$

where RNB represents the target value output of the naive Bayes classifier.

5. EVALUATION RESULTS AND DISCUSSIONS

Proposed approach on two indoor scene classification datasets were evaluated. One video for each indoor scene (corridor, staircase, and room) and a total of three videos (2.4 GHZ, 1.5 m to 3 m) with resolution of 1280 × 720 pixels obtained from parrot AR drone MAV are considered and the image frames of the indoor videos are included for analysis in Dataset-2. Proposed scene recognition algorithm is mainly developed to recognise three indoor scenes i.e., corridor, staircase, and room, but not to discriminate between intra-class (e.g. living room *vs.* bedroom) type of indoor scenes. Moreover, in this work some basic categories of scenes (bathroom, bedroom, children room, class room, computer room, dining room, game room, hospital room, living room, locker room, meeting room, operating room and waiting room) can be grouped and considered belonging to a single category (e.g. room). In the SVM classifier, a 'linear' kernel function is used. For the kNN classifier, the Euclidean distance was selected as the distance metric. In the Naïve bayes classifier, kernel density estimation method (KDE) is used. The classification results for GIST, CENTRIST, HODMG, LBP and enhanced GIST features using SVM, Naïve bayes and k-NN classifiers for Dataset-1 and Dataset-2 are presented in Tables 1, 3, and 4. From Table 1, it can be observed that proposed enhanced GIST descriptors using SVM classifier gives better classification accuracy compared to the other features. Properties of input videos are listed in Table 2. All the experiments were carried out on the ground station on laptop computer with an Intel CPU operating at 2.20 GHz and 2 GB RAM. The class-wise classification accuracies of the two indoor datasets for the proposed method (enhanced GIST descriptors + SVM) are as shown in the form of confusion matrices in Fig. 9. Dataset-1 has 72 test images (24 images per each category), out of which totally 2 'staircase' images were wrongly classified as 'corridor' and 'room'. Dataset-2 has 150 test images (50 images per each category), among which one of the 'staircase' images were misclassified as 'room'. The strong diagonal in all confusion matrices indicates better classification performance. From the confusion matrix, it is inferred that confusion happens between staircase and corridor, staircase and room. Maximum recognition rates of 97.22 per cent and 99.33 per cent is obtained using SVM classifier for dataset-1 and dataset-2, respectively. From the results, it is inferred that the SVM is a more suitable and effective classifier for the classification of indoor scenes. Table 5. Compares the effectiveness of the SVM classifier to the k-Nearest Neighbour classifier and Naïve Bayes classifier. The results based on SVM classifier, when compared with kNN and Naïve Bayes classifier shows higher recognition accuracy. The average computation time taken per frame is shown in Table 5. In Table 5, linear SVM performs better compared to kNN and NBC because the data is linearly separable using SVM and less prone to overfitting than KNN and NBC classifiers.

Table 1. Classification results of the SVM and Naïve Bayes classifiers for Dataset-1 and Dataset-2

Classifier	Types of features	Dataset-1	Dataset-2
		Accuracy (%)	Accuracy (%)
SVM Classifier	Enhanced GIST	97.22	99.33
	GIST	76.38	72.66
	CENTRIST (Not using PCA)	69.44	53.33
	HODMG	54.16	45.33
	LBP8,1 u2	58.33	Not Converge
	LBP8,1 riu2	58.33	45.33
	LBP8,2 u2	Not Converge	59.33
	LBP8,2 riu2	59.72	59.33
	LBP16,2 riu2	70.83	65.33
Naïve bayes classifier	Enhanced GIST	76.38	50.66
	GIST	87.50	74.66
	CENTRIST (Not using PCA)	59.72	44
	HODMG	65.27	37.33
	LBP8,1 u2	59.15	51.33
	LBP8,1 riu2	52.77	38.66
	LBP8,2 u2	61.11	64
	LBP8,2 riu2	48.61	58
	LBP16,2 riu2	48.61	58.66

Table 2. Properties of input videos

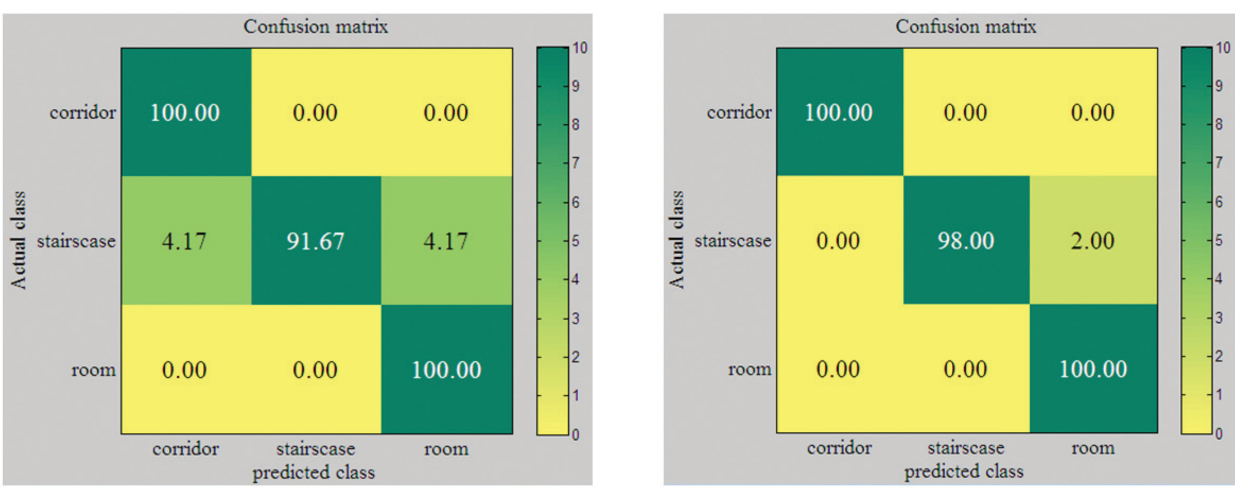
Properties	Video 1 (Corridor)	Video 2 (Staircase)	Video 3 (Room)
No. of frames	2252	2444	760
Frame rate	30 fps	30 fps	30 fps
Resolution in pixels	1280×720	1280×720	1280×720

5. CONCLUSIONS

Indoor scene recognition algorithm based on combining Histogram of horizontal and vertical directional morphological gradient features and GIST features is presented in this paper. Authors have evaluated the five image descriptors, namely GIST descriptors, CENTRIST (Not using PCA), histogram of directional morphological gradient features, local binary pattern features and enhanced-GIST descriptors on two different indoor datasets. The combination of local features (HODMG) and global image descriptor (GIST) has proven to provide good results for indoor scene recognition task. Furthermore, the SVM incorporating enhanced-GIST descriptors obtain the best performance for classifying the corridor, staircase and room type of indoor scenes, with a recognition rates of 97.22 per cent for dataset-1 and 99.33 per cent for dataset 2, compared to kNN and Naïve Bayes classifiers. Our experiments also demonstrate that the proposed method can be used as a scene classifier for MAV navigation, and Enhanced-GIST descriptors gives better results when compared with the other individual state-of-the-art descriptors such as GIST, CENTRIST (Not using PCA) and LBP.

REFERENCES

1. Quattoni, A. & Torralba, A. Recognizing indoor scenes. *In* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, 2009. doi: 10.1109/CVPR.2009.5206537.
2. Oliva, A. & Torralba, A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comp. Vis.*, 2001, **42**(3), 145-175. doi: 10.1023/A:1011139631724.
3. Oliva, A. & Torralba, A. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.*, 2006, **55**, 23-36.



(a) (b)

Figure 9. Confusion matrices obtained for indoor datasets: (a) Dataset-1 and (b) Dataset-2.

Table 3. Classification results of the K-NN classifier for Dataset-1

NN classifiers with different k values	Number of training samples for Dataset-1 and their accuracy in percentage (for 72 test images) with 180 samples								
	Enhanced GIST	GIST	CENTRIST not using PCA	HODMG	LBP8,1 u2	LBP 8,1 riu2	LBP8,2 u2	LBP8,2 riu2	LBP16,2 riu2
K=1	38.88	63.88	51.38	38.88	45.83	61.11	56.94	47.22	47.22
K=2	38.88	63.88	51.38	38.88	45.83	61.11	56.94	47.22	47.22
K=3	37.50	70.83	54.16	37.50	52.77	61.11	52.77	47.22	45.83
K=4	44.44	69.44	58.33	44.44	54.16	65.27	54.16	50	48.61
K=5	44.44	69.44	59.72	44.44	61.11	63.88	62.50	44.44	44.44
K=6	51.38	68.05	56.94	51.38	56.94	55.55	51.38	43.05	50
K=7	50	70.83	59.72	50	58.33	55.55	55.55	45.83	51.38
K=8	51.38	69.44	56.94	51.38	59.72	55.55	50	47.22	50
K=9	54.16	70.83	58.33	54.16	62.50	51.38	48.61	45.83	51.38
K=10	50	68.05	59.72	50	59.72	50	51.38	48.61	47.22
K=11	58.33	68.05	58.33	58.33	54.16	51.38	52.77	48.61	45.83
K=12	54.16	68.05	59.72	54.16	54.16	55.55	48.61	50	43.05
K=13	54.16	68.05	58.33	54.16	58.33	54.16	52.77	45.83	48.61
K=14	50	66.66	55.55	50	58.33	52.77	54.16	47.22	50
K=15	52.77	66.66	56.94	52.77	61.11	51.38	54.16	48.61	43.05
K=16	52.77	68.05	54.16	52.77	61.11	50	54.16	47.22	47.22
K=17	55.55	68.05	58.33	55.55	58.33	44.44	56.94	47.22	48.61
K=18	52.77	68.05	55.55	52.77	59.72	45.83	54.16	47.22	55.55
K=19	55.55	68.05	59.72	55.55	61.11	54.16	56.94	50	54.16
K=20	58.33	68.05	58.33	58.33	63.88	54.16	61.11	51.38	58.33
K=21	58.33	66.66	58.33	58.33	61.11	56.94	56.94	52.77	52.77
K=22	58.33	68.05	56.94	58.33	61.11	59.72	61.11	51.38	55.55
K=23	58.33	68.05	58.33	58.33	61.11	55.55	55.55	44.44	48.61
K=24	55.55	66.66	56.94	55.55	59.72	56.94	55.55	45.83	52.77
K=25	58.33	66.66	55.55	58.33	59.72	54.16	54.16	52.77	50

doi:10.1016/S0079-6123(06)55002-2

4. Wu, J. & Rehg, J.M. CENTRIST: A visual descriptor for scene categorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**(8), 1489-1501. doi: 10.1109/TPAMI.2010.224
5. Ojala, T.; Pietikäinen, M. & Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**(7), 971–987. doi: 10.1109/TPAMI.2002.1017623
6. Vailaya, A.; Jain, A. & Zhang, H.J. On image classification: city images vs. Landscapes. *Pattern Recognit*, 1998, **31**, 1921-1935. doi:10.1109/IVL.1998.694464
7. Espinace, P.; Kollar, T.; Roy, N. & Soto, A. Indoor scene recognition by a mobile robot through adaptive object detection. *Robotics Autonomous Sys.,* 2013, **61**(9), 932-947. doi: 10.1016/j.robot.2013.05.002
8. Cakir, F.; Güdükbay, U. & Ulusoy, Ö. Nearest-neighbour based metric functions for indoor scene recognition. *Comput. Vis. Image Underst.,* 2011, **115**(11), 1483-1492. doi: 10. 1016 /j. cviu. 2011. 07. 007.
9. Fornoni, M. & Caputo, B. Indoor scene recognition using task and saliency driven feature pooling. *In* Proceedings of the British Machine Vision Conference, Surrey, UK, 2012. doi: 10.5244/C.26.98
10. Yu, J.; Hong, C.; Tao, D. & Wang, M. Semantic embedding for indoor scene recognition by weighted hypergraph learning. *Signal Processing*, 2015, **112**, 129-136. doi: 10.1016/j.sigpro.2014.07.027

Table 4. Classification results of the K-NN classifier for Dataset-2

NN classifiers with different k values	Number of training samples for Dataset-2 and their accuracy in percentage (for 150 test images) with 300 samples								
	Enhanced GIST	GIST	CENTRIST not using PCA	HODMG	LBP 8,1 u2	LBP 8,1 riu2	LBP 8,2 u2	LBP 8,2 riu2	LBP 16,2 riu2
K=1	38	88.66	41.33	38	46.66	52	40	41.33	48.66
K=2	38	88.66	41.33	38	46.66	52	40	41.33	48.66
K=3	38	84.66	21.33	38	30.66	54.66	32.00	39.33	42.66
K=4	44.00	82.66	25.33	44.00	35.33	56.00	24	40.66	42.66
K=5	22.00	76.66	22.00	22.00	37.33	52.66	28.00	44.66	43.33
K=6	38.66	79.33	27.33	38.66	30.66	51.33	33.33	48	42.66
K=7	30.00	78	29.33	30.00	30.00	51.33	30.66	34.66	45.33
K=8	37.33	80	27.33	37.33	31.33	42.66	42.00	38.66	49.33
K=9	38	76.66	34.66	38	34	49.33	40	39.33	48
K=10	39.33	72.66	40.66	39.33	52	46.66	42.66	39.33	49.33
K=11	39.33	74.66	35.33	39.33	41.33	47.33	42.66	40.66	47.33
K=12	39.33	74	43.33	39.33	47.33	48	52.66	41.33	47.33
K=13	41.33	74	43.33	41.33	42.00	46.66	48	39.33	47.33
K=14	39.33	75.33	50.66	39.33	52	48	52	42.00	48
K=15	42.66	74.66	53.33	42.66	46.66	46	54.66	42.00	48
K=16	40	74	57.33	40	54	52	52	42.00	50
K=17	41.33	74.66	53.33	41.33	54.66	51.33	51.33	46	51.33
K=18	43.33	74.66	54.66	43.33	60.66	58.00	53.33	44.00	54.66
K=19	47.33	71.33	55.33	47.33	59.33	56.00	52.66	46.66	55.33
K=20	56.00	72.66	59.33	56.00	58.66	54.66	52.66	44.66	54.66
K=21	56.00	72	60	56.00	56.00	53.33	52	44.66	53.33
K=22	46.66	72	62.66	46.66	58.66	52.66	62	46	57.33
K=23	48.66	72.66	64	48.66	57.33	51.33	63.33	46	57.33
K=24	43.33	71.33	64.66	43.33	60	52.66	61.33	46	57.33
K=25	46.66	71.33	64.66	46.66	59.33	52.66	61.33	45.33	48.66

Table 5. Comparison of SVM, kNN and Naïve Bayes classifier

Type of Feature	Classifier method	Dataset-1		Dataset-2	
		Recognition rate (%)	Average time elapsed, in seconds per frame	Recognition rate (%)	Average time elapsed, in seconds per frame
Enhanced-GIST	SVM	97.22	1.60	99.33	1.91
	kNN	58.33	1.22	56.00	1.18
	NBC	76.38	20.68	50.66	21.20

11. Khan, S.; Hayat, M.; Bennamoun, M.; Sohel, F. & Togneri, R.A discriminative representation of convolutional features for indoor scene recognition. *IEEE Trans. Image Process.,* 2016, **25**(7), 3372-3383. doi: 10.1109/TIP.2016.2567076.
12. Anbarasu, B.; Amardeep, A. & Anitha, G. Obstacle detection and collision avoidance for indoor navigation of a MAV using Sonar Sensors. *Int. J. Appl. Eng. Res.,* 2015, **10**(8), 6511-6514.
13. Haralick, R.M.; Sternberg, S.R. & Zhuang. Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1987, **9**(4), 532-550. doi: 10.1109/ TPAMI. 1987. 4767941
14. Soille, P. Morphological image analysis: Principles and applications. Springer-Verlag, 2003.
15. Cortes, C. & Vladimir, V. Support-vector networks. *Mach. Learn.*, 1995, **20**(3), 273-297. doi: 10.1023/A:1022627411411

16. Pontil, M. & Verri, A. Support vector machines for 3D object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**(6), 637-646.
doi: 10.1109/34.683777
17. Fukunaga, K. Introduction to statistical pattern recognition. Academic Press, 1990.
18. Cover, T.M. & Hart, P.E. Nearest neighbour pattern classification. *IEEE Trans. Inform. Theory*, 1967, **13**(1), 21-27.
doi: 10.1109/TIT.1967.1053964
19. Mitchell, T. Machine Learning. McGraw Hill, 1997.sss
20. John, G.H. & Langley, P. Estimating continuous distributions in Bayesian classifiers. *In* Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence, San Francisco, 1995.

ACKNOWLEDGEMENT

The authors acknowledge the support rendered by Division of Avionics, Department of Aerospace Engineering, Madras Institute of Technology, Anna University, Chennai.

CONTRIBUTORS

Mr B. Anbarasu has obtained his BE (ECE) from Madurai Kamaraj University and ME (Avionics) from Anna University. Currently he is working as Teaching Fellow in the Department of Aerospace Engineering at MIT Campus, Anna University, Chennai and pursuing his research in indoor navigation. His areas of interest are: Avionics, UAV systems and Image processing.

He has contributed in the design and development of the indoor scene recognition algorithm and interpretation of results.

Dr G. Anitha has obtained her BE (EEE) from ACCET, Karaikudi and MTech (Control & Instrumentation) from IITM, Chennai and PhD in Avionics from Anna University, Chennai. She is working as Assistant Professor (Senior Grade) in the Division of Avionics, Department of Aerospace Engineering, MIT, Anna University, Chennai. Her areas of interest are: Avionics, aircraft navigation, guidance, and control systems and image processing.

She has contributed in design and development of the research methodology and was responsible for overall guidance in implementation of the proposed algorithm.