International Journal of Computer Trends and Technology (IJCTT) – volume 23 Number 3 – May 2015

A Novel Approach for Classification of Indoor Scenes

Gagandeep Kaur[#1], Dr. Amandeep Verma[*2]

[#] *Department of Computer Science, Punjabi University, India*

* *Department of Computer Science, Punjabi University,India*

***Abstract*— Scene classification is recently growing area of research in computer vision. A variety of approaches has been proposed for scene classification. The literature addresses the issues involved in indoor scene classification. The segmentation based approaches suffer from poor performance of segmentation and object-based approaches involve series of complex tasks like segmentation, training a large number of classifier and recognition. In this study, a novel approach for classification of indoor scenes into multiple classes has been proposed. The proposed feature representation is entirely based on extracting structural properties of the scene images. The proposed method uses Gaussian filter in pre-processing phase to reduce noise from image followed by using morphological operations to extract edge features from image. The one-vs-all Support Vector Machines (SVM) learning model is employed for learning and classification. To test the performance of classification system, a database of five indoor classes i.e. bedroom, living room, dining room, office and kitchen has been taken from MIT-indoor dataset. The images have been taken under different under different illumination conditions and different viewpoints. The accuracy of 84% and sensitivity of 56% has been obtained for five indoor classes..**

***Keywords*— Indoor Scene Classification, Structural properties, Morphological Gradient.**

I. INTRODUCTION

To design and manufacture robotic replicates of humans is the fundamental problem of computer vision. The ability of robot to recognize the environment around it, is the principal foundation of complexity of works assigned to mobile robot. The scene classification also known as scene recognition or scene understanding is probably one of the most basic and difficult problem of machine vision which enables the machine to perceive the scene images the way human do. Further content based image retrieval and indexing also utilizes similarity based image classification techniques[1].

A plethora of research has been done in the field of scene classification, but mainly for outdoor scene classification. The classification rates of state-of-art methods [2][3][4] for outdoor scenes have been found to be above 70%. But when same techniques are tested on indoor datasets like MIT-indoor dataset[5], KDH-IDOL dataset, the performance drops dramatically (□□ 40%).There are several difficulties in handling indoor scenes. First of all there is no significant difference in color and texture of indoor scenes. The second reason is that there is no proper spatial layout of scene. The third reason is that different view-points of indoor environment give different visual appearance. The recent methods based on segmentation[6] or objects recognition[7] demand strong geometric constraints of the objects of the scene, which are less application in scene classification as geometric appearance of constituent objects changes with change in different spatial configurations within the scene.

It is believed that the appropriate representation of the features is a key to the success of scene classification[8]. The main source of controversy has been between using the global features that captures the holistic representation[3] of the images versus local features like identity of constituent objects[9]. In terms of classification, as noticed before [3][8] the holistic representation of the scene image provide important clues about identity of scene.

Following the motivation above, the holistic representation of image has been used. The proposed method is completely based on capturing rough geometry of the scene. The idea is to capture background features of the image which have high probability of being invariant with respect to different colored objects and different number of objects. These structural properties of the scene like rectangles, flat surfaces, and vertical surfaces provide important clues about identity of the scene. The hypothesis of the proposed method is that color or texture information does not provide significant discriminative features in case of indoor scenes. The vertical edge, horizontal edges etc. properties of the image gives strong clues about the structure of the scene.

The features used for indoor scene classification is based on extracting the rough geometric spatial structures of indoor scene images which are quite helpful in recognizing the scene. A training set of representative images for each of the five indoor classes has been chosen for this learning based classification system. The supervised learning based approach has been adopted by using manually annotated training images. The parameters learnt during training help the system to classify an unseen image into one of the pre-defined categories. The performance of system is determined on the basis of parameters precision, sensitivity and accuracy.

The paper has been organized as follows: section II discusses the related work in classification of indoor scenes, section III gives details of design and implementation of proposed method, section IV discusses the results and the last section V presents conclusion and future work.

II. RELATED WORK

There is a large array of literature associated with outdoor scene classification. These techniques utilize color and texture dissimilarities of natural scenes to categorize the scene images[10][11]. The most popular techniques are GIST[3], Spatial Pyramid Matching[4] and Bag-of-Words[1]. The approaches used for indoor scene classification usually consider the importance of identities of the constituent objects or local features.

There are three main approaches used for indoor scene classification: The first one is inspired from traditional scene classification methods[3][12] which uses either local or global features to represent an image. The combination of global and local features is proposed in for classification of indoor scenes[5]. The global features are computed using GIST[3] while manually annotated regions of interest (ROIs) are used as local features. The visual vocabulary was learnt for each ROIs using SIFT[13] descriptors followed by k-means clustering. The HSV-SIFT[14] descriptor integrated with GIST further improves the performance by providing complementary information. The recent techniques also utilize 3D geometry[15] of the scene along with 2D global GIST features. Recently a new visual descriptor CENTRIST[16] has been proposed which retains the global as well as local properties of the scene image. An alternate approach[17] utilizes orientations features of constituent parts for indoor scene classification.

The second approach is based on segmentation. The performance of segmentation based classification is entirely based on efficiency of segmentation algorithm. The features like size, shape, planarity etc. of segmented parts are used to train region classifiers[6][18]. The main difficulty in segmentation based approach is unavailability of efficient segmentation techniques that can segment out cluttered indoor scenes into its constituent parts.

Another approach which is based on object recognition[7] involves the series of processes like segmentation, feature grouping, recognition etc. The complexity of the processes involved and amplifying error along stream of processing hamper the performance of the object-based scene classification methods. One of the recent techniques is based on using object-filters[19] for classification of scene images. This approach uses a large number of pre-trained object-filters which run on images and generalized response of object-filters is used as representation of the image.

III. THE APPROACH

The structural similarities and dissimilarities have been used to classify the scene images. This simplified approach of feature extraction bypasses segmentation and recognition of objects to classify the scene. The algorithm consists of four phases: Dataset Preparation, Pre-processing, Feature extraction and classification. The detailed flow-chart of the algorithm has been shown in Figure 1 with the help of flow-chart.

A. Dataset Preparation

The MIT-indoor dataset[20] is a standard benchmark for indoor scene classification. It contains a total of 15620 images for 67-indoor scenes like bedroom, living room, inside mall, conference hall, bakery etc. The pictures in the dataset were gathered from distinctive sources: online picture inquiry devices (Google and Altavista), online photograph offering locales (Flickr) and the Label Me dataset. The images have been taken under different illumination condition and different viewpoints.

The five indoor classes namely bedroom, corridor, dining room, kitchen and living room have been chosen from MIT-indoor dataset to evaluate the performance of proposed algorithm. The thirty five images of each category have been randomly selected which are further divided into disjoint training set and testing set. The training set contains 125 images (25 images for each category) and testing set contains 50 images (10 images for each category).

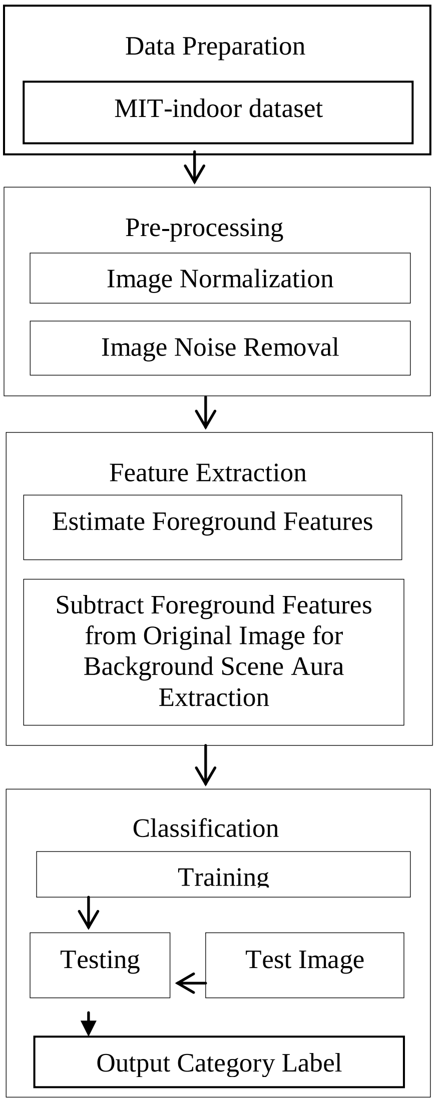

Fig. 1: Flow chart of proposed algorithm

B. Pre-processing

The pre-processing phase involves a series of steps: The first step is to convert RGB image into gray-scale image. The gray scale images have been used in the entire process. The next step is to resize the images. All the images are resized to 300×400. The last step involves image smoothing to reduce the noise. The Gaussian filter[21] has been to blur the images and reduce the noise. The smoothing effect depends upon the standard deviation sigma (σ). The Gaussian filter G is given by equation 3.1.

$$G_g(n_1, n_2) = e^{-(n_1^2+n_2^2)2\sigma^2}$$

---Equation 3.1

.

C. Feature Extraction

The proposed algorithm for indoor scene classification extracts the rough geometrical features of the scene ignoring color and detailed textural features. The morphological gradient has been used to extract structural properties of the image. The mathematical morphology is a powerful tool for extracting geometric features. It finds the prominent candidate components of the scene and suppresses the detailed textural information. The morphological dilation of Image I with structuring element H is given below by equation 3.2.

$$[I \oplus H](u,v) = \max_{(i,j)\in H}\{I(u-i, v-j) + H(i,j)\}$$

---Equation 3.2

The morphological dilation of image makes bright regions surrounded by dark regions to grow in size and dark regions surrounded by bright region to shrink in size. The effect is most marked at places where intensity value changes rapidly whereas uniform regions remain almost unchanged. The undesired effect of the dilation is distortion of boundaries of objects, So Morphological closing is done to fill the boundaries of the objects. The resultant image highlights the prominent features of the image, so these are named as foreground features.

The resultant foreground features are subtracted from original gray scale image to obtain morphological gradient which represents the structural properties of the scene. This background scene information ignores the color and texture information of foreground objects. Figure 2 shows that the rough geometrical features of image indicate that given image is a bedroom image. The proposed algorithm avoids the direct interpretation of the identities of the objects to classify indoor scenes.

Fig. 2: An example of bedroom image taken from MIT-indoor dataset and its corresponding morphological gradient.

D. Classification

The discriminative SVM model has been used for learning and classification. One-vs-all SVM[22] is used for multi class classification. The five SVM models have been trained for classification of five classes. The decision is based on the distance of test from each hyper-plane and final output is the class corresponding to the hyper-plane for which the distance is largest. The training set of total 125 images has been used to train the One-vs-all SVMs and the trained model assigns a category label to the test image.

The following is the algorithm for pre-processing and background feature extraction.

Algorithm 1: Preprocessing and feature extraction algorithm

I.	Load the 3-D (colored) test image as test object matrix T_m
II.	Convert the test image matrix to grayscale image matrix G_m
III.	Define the Gaussian Filter G_f
IV.	Apply Gaussian Filter G_f on G_m to produce the de-noised G_{md}
V.	Define the dilation object of adequate shape and size SE
VI.	Dilate the image G_{md} with respect to SE to produce the image object A_I
VII.	Perform morphological closing of the image object A_I to obtain foreground features F_{EF}
VIII.	Subtract the F_{EF} from G_m to produce morphological gradient F(BG)
IX.	Return the F(BG)

IV. EXPERIMENTAL RESULTS

In this section, we first provide detailed analysis on the model and parameters with the MIT Indoor-67 dataset. The MATLAB environment has been used to implement the proposed algorithm. The proposed algorithm has been implemented using basic MATLAB programming. The performance of the system is evaluated using performance parameters true positive (TP), true negative (TN), false positive (FP), false negative (FN), Precision, Sensitivity and accuracy. Five indoor scene classes namely bedroom, corridor, dining room, kitchen and living room have been chosen from MIT-indoor dataset.

The results have been obtained for individual class. The Table 1 shows the confusion table for the five classes i.e. bedroom, corridor, dining room, kitchen and living room. The results show that the performance of the proposed method varies significantly for different classes. The proposed algorithm is highly efficient in capturing geometric features in bedroom class where the prominent objects are windows, bed, lamp etc. However the poor results in case of kitchen is due to more uniform appearance of wooden drawers and shelves, where the algorithm fails to capture prominent features. The most confusing classes are kitchen and living room. Due to high intra-class variability, the results are poor for living room.

Table 1: Confusion Table for five indoor classification

	Bedroom	Corridor	Dining Room	Kitchen	Living Room
Bedroom	8	0	1	0	1
Corridor	0	7	0	3	0
Dining Room	1	0	6	3	0
Kitchen	1	0	1	3	5
Living Room	2	1	2	0	4

The performance of the proposed algorithm for each indoor class is shown in Table 2. The performance parameters illustrate the performance of proposed algorithm for each scene class. The overall results are best for corridor which is the class with least intra-class variability. There is significant intra-class variability but still our algorithm shows good results for this class too by capturing important structural features.

Table 2: Overall performance of proposed classification model

	TP	FP	TN	FN	Accuracy	Sensitivity
Bedroom	8	4	36	2	88%	80%
Corridor	7	1	39	3	92%	70%
Dining Room	6	4	36	4	84%	60%
Kitchen	3	6	34	7	78%	30%
Living Room	4	5	35	6	78%	40%
Average					84%	56%

The graphical representation of three performance parameters precision, recall and accuracy has been shown in Figure 3. The graph shows five classes bedroom, corridor, dining room, kitchen and living room with their respective performance graphs. The graphs show highest accuracy and highest sensitivity for corridor while highest sensitivity for bedroom class. The average accuracy of 84% has been achieved for these five indoor classes. The average value of precision, sensitivity and accuracy of proposed model has also been to evaluate the overall performance of proposed method.

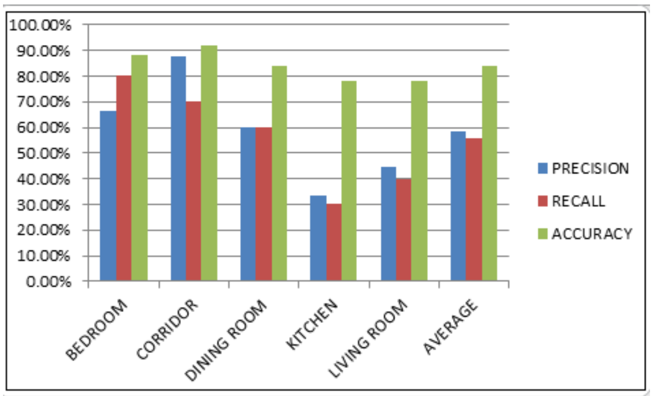

Figure. 3 : Performance graph of proposed method for five classes: bedroom, corridor, dining room, kitchen and living room.

V. CONCLUSIONS

Keeping in view the importance of scene classification and difficulties in classification of indoor scenes, this research work presents some measures at feature extraction phase. The rough geometry for different indoor classes has been used as discriminative feature for classification. This completely supervised learning algorithm for scene classification is tested on five indoor classes i.e. bedroom, living room, dining room, kitchen and corridor. The obtained results indicate that the proposed feature extraction method is highly efficient for recognizing bedroom and corridor. Although it shows poor performance on mostly uniform class i.e. kitchen and highly cluttered living room, but the overall performance shows that idea of extracting structural properties of scenes for indoor scene classification can help to improve the accuracy of indoor scene classification systems.

ACKNOWLEDGMENT

I owe a great thanks to Dr. Amandeep Verma, my research guide, for all help and guidance provided by him throughout the study. I really appreciate all the valuable time he has spent in helping me and discussing various problems on the way. It has been very fortunate to have him as my thesis guide. I express my sincere gratitude towards him.

I would also like to thank Mrs. Nivit Gill and Mrs. Amardeep Kaur, Assistant Professor in the Department of Computer Science, Punjabi University Regional Center for Information Technology and Management, Mohali for their moral support. I am highly indebted to them for clarifying my doubts and for providing suggestions and positive criticism on my work.

REFERENCES

[1]C.-F. Tsai, “Bag-of-Words Representation in Image Annotation: A Review,” *ISRN Artif. Intell.*, vol. 2012, pp. 1–19, 2012.

[2]J. Vogel and B. Schiele, “Semantic Modeling of Natural Scenes for Content-Based Image Retrieval,” *Int. J. Comput. Vis.*, vol. 72, no. 2, pp. 133–157, Jul. 2006.

[3]A. Oliva, W. Hospital, and L. Ave, “Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope ,” *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[4]S. Lazebnik, C. Schmid, and J. Ponce, “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 2 (CVPR’06)E*, vol. 2, pp. 2169–2178, 2006.

[5]A. Quattoni, U. C. B. Eecs, and A. Torralba, “Recognizing Indoor Scenes,” *IEEE CS Conf. Comput. Vis. Pattern Recognit.*, 2009.

[6]S. Gupta, P. Arbelaez, and J. Malik, “Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images,” *2013 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 564–571, Jun. 2013.

[7]P. Espinace, T. Kollar, N. Roy, and a. Soto, “Indoor scene recognition by a mobile robot through adaptive object detection,” *Rob. Auton. Syst.*, vol. 61, no. 9, pp. 932–947, Sep. 2013.

[8]J. Wu and J. M. Rehg, “CENTRIST: A Visual Descriptor for Scene Categorization.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, Dec. 2010.

[9]L. Antanas, M. Hoffmann, P. Frasconi, T. Tuytelaars, and L. De Raedt, “A relational kernel-based approach to scene classification,” *2013 IEEE Work. Appl. Comput. Vis.*, pp. 133–139, Jan. 2013.

[10]C. S. Chan, “A fuzzy qualitative approach for scene classification,” *2012 IEEE Int. Conf. Fuzzy Syst.*, vol. 1, pp. 1–8, Jun. 2012.

[11]G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, and D. Maupertuis, “Visual Categorization with Bags of Keypoints,” *Work. Stat. Learn. Comput. Vision, Eur. Conf. Comput. Vis.*, vol. 1, 2004.

[12]D. Gokalp and S. Aksoy, “Scene Classification Using Bag-of-Regions Representations,” *2007 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, Jun. 2007.

[13]D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[14]H. Madokoro, a. Yamanashi, and K. Sato, “Unsupervised semantic indoor scene classification for robot vision based on context of features using Gist and HSV-SIFT,” *Pattern Recognit. Phys.*, vol. 1, no. 1, pp. 93–103, Aug. 2013.

[15]A. Swadzba and S. Wachsmuth, “Indoor Scene Classification using combined 3D and Gist Features,” *Comput. Vision–ACCV , Springer Berlin Heidelberg, Springer Berlin Heidelberg*, vol. 6493, pp. 201–215, 2011.

[16]J. Wu and J. M. Rehg, “CENTRIST: A visual descriptor for scene categorization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, 2011.

[17]L. Xie, J. Wang, B. Guo, B. Zhang, and Q. Tian, “Orientational Pyramid Matching for Recognizing Indoor Scenes,” *2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3734–3741, Jun. 2014.

[18]L. Nan, K. Xie, and A. Sharf, “A search-classify approach for cluttered indoor scene understanding,” *ACM Trans. Graph.*, vol. 31, no. 6, p. 1, Nov. 2012.

[19]L. J. Li, H. Su, Y. Lim, and L. Fei-Fei, “Object bank: An object-level image representation for high-level visual recognition,” *Int. J. Comput. Vis.*, vol. 107, no. 1, pp. 20–39, 2014.

[20]Ariadna Quattoni and Antoni Torralba, “Indoor scene database,” 2009. [Online]. Available: http://web.mit.edu/torralba/www/indoor.html.

[21]R. C. W. Gonzalez, *Digital Image Processing*. Pearsons Eucation India, 3rd ed., 2009.

[22]A. Pronobis, B. Caputo, P. Jensfelt, and H. I. Christensen, “A realistic benchmark for visual indoor place recognition,” *Rob. Auton. Syst.*, vol. 58, no. 1, pp. 81–96, Jan. 2010.