IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ADAPTIVE INDOOR SCENE CLASSIFICATION WITH MULTI-SVM CLASSIFICATION TO SOLVE MULTI-CLASS PROBLEM

Japneet Kaur*, Rimanpal Kaur
* Department of Computer Science CGC Technical Campus Jhanjeri,Mohali, India

DOI: 10.5281/zenodo.52502

ABSTRACT

In recent years, indoor scene recognition has attracted much attention and its research has rapidly expanded by not just engineers but also neuroscientists, since it has many potential applications in computer vision communication and automatic access control system. Especially, indoor scene recognition is an important part of computer vision and feature recognition as the first step of automatic uninterrupted robotic movement or computer vision applications like automatic interior designing algorithms. However, the indoor scene detection is not straightforward because it has lots of variations of image appearance, such as light effect, occlusion, image orientation, illuminating condition and object variety. Many novel methods have been advanced to resolve each variation listed above. For example, the template-matching methods are used for indoor scene localization and detection by computing the correlation of an input image to a standard and training scene appearance or pattern. The feature invariant approaches are used for feature detection of bed, chair, cabinet, table, door, electrical or electronic items, etc. The appearance-based methods are used for indoor feature detection with support vector machine and information theoretical approach. Nevertheless, implementing the methods altogether is still a difficult task. Fortunately, the images used in this project have some degree of uniformity thus the detection algorithm can be easy: first, the all the faces are vertical and have frontal view; second, they are under almost the same illuminate state. This project presents an indoor scene detection technique mainly based on the appearance based feature segmentation, SVM training and SVM classification methods to recognize the indoor scenes.

KEYWORDS: Multi-SVM, support vector machine, multi-class problem, indoor classification.

INTRODUCTION

Scene classification is aimed at labeling an image into semantic categories (room, office, mountain etc). It is an important task to classify, organize and understand thousands of images efficiently[1]. From application point of view, scene classification is useful in-content based image retrieval. As accurate classification of an image helps in better organization and browsing[3,4]. Scene classification is highly valuable in remote navigation also.
Some of the feature descriptors to be used are:-

- **GIST-** A typical GIST is computer over a complete image for the scene classification task. It falls in the global image descriptor category.
- **SIFT-** The typical use of SIFT is to match the local regions in two images on the basis of their reconstruction, alignment or other similar. SIFT can be used for the purpose of identification of some specific objects by using BW (Bag of Words) model.
- **HOG-** Histogram of Oriented Gradient (HOG) [21] is used for object-recognition. It is based on computing edge- gradients. Typical HOG works in the sliding window fashion for object detection applications. HOG computes the complete image after dividing it into the smaller cells, called blocks. HOG can be used alongside SVM for feature detection using classification.
- **CENTRIST-** CENTRIST (Census Transform histogram) is a novel visual descriptor, which is more robust to illumination changes, gamma variation etc. as compared to GIST [15] and SIFT [16]. CENTRIST is histogram of Census Transform (CT) values. CT compares intensity value of a pixel with its neighboring pixels and assigns value 1 or 0 to those pixels. After that the decimal number corresponding to this sequence

of 8 neighboring binary digits is computed and used as CT value of center pixel. This descriptor retains the local as well as global structure of the scene. However, there are several limitations of this descriptor. It is not invariant to rotation and scale changes. It also does not consider color information. Further it cannot be used for precise shape description.

EXPERIMENTAL DESIGN

The proposed system for the scene classification consists of the various techniques such as flexible feature descriptor, feature optimization and classification algorithms. The proposed model flow has been described in the following flow table 3.1.

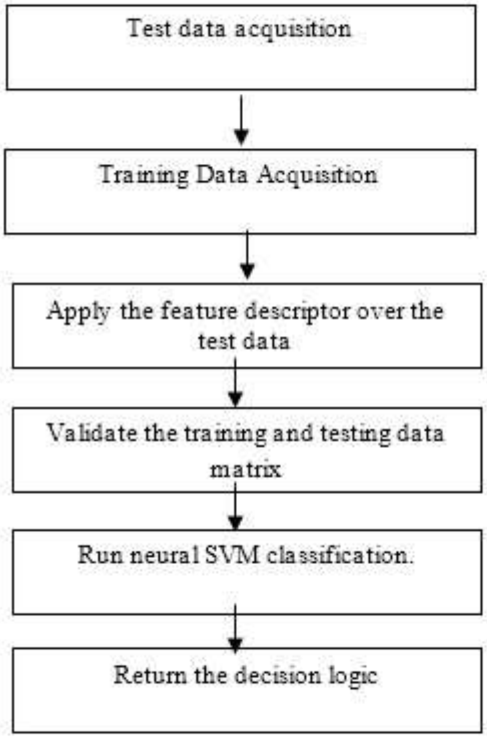
Test data acquisition
Training Data Acquisition
Apply the feature descriptor over the test data
Validate the training and testing data matrix
Run neural SVM classification.
Return the decision logic

Figure 2.1: The workflow diagram describing the proposed model

During the very first phase of the proposed model, the main object is to properly describe the feature descriptor matrix for the training and testing data. The first phase involves all of the practices related to the data preparation for the training and the testing data matrices. The experiment for the proposed model starts with the data collection; hence this phase plays the vital role in the proposed model.

The popular indoor database constructed and collected by the researched at MIT institute has been utilized for the testing of the proposed model performance. The MIT dataset contains the 15620 image, which includes the data of various indoor categories. The data of total 67 categories has been added to the training matrix.

The preprocessing phase includes the various practices to mitigate the image noise, image size neutralization, feature vector resizing or reshaping, etc.

1. The first step in the preprocessing involves the acquisition of the image matrix by using the acquisitions functions pre-programmed in the simulation environment.
2. All of the images in the dataset are of the different sizes. For the data classification the image data must be of the similar size, which requires all of the images to be resized on to the same size by using the image resizing methods.
3. Converting the image to the grayscale isthe another step involved in the pre-processing module. The grayscale conversion converts the image matrix to the 2-D matrix, which is utilized to obtain the low level features from the data matrix.

4. The Gaussian filter is utilized to de-noise the image data from the contiguous noise produced due to the blurring effects, signaling interference, etc. The Gaussian filter is created and applied over each and every image in the dataset before the application of feature normalization or the feature classification module.
5. The feature representation and the feature resizing involve the final step before the classification algorithm. The feature resizing is done according to the size of the input data matrix and the training matrix column size. Any of the variance is covered by resizing or trimming the feature vector. The feature representation involves the set of calculations over the final feature descriptor to churn out the meaningful features by using the methods such as TF-IDF, Standard deviation, 1-D or 2-D mean factor, 1-2 or 2-D median factor over the input feature descriptor vector.

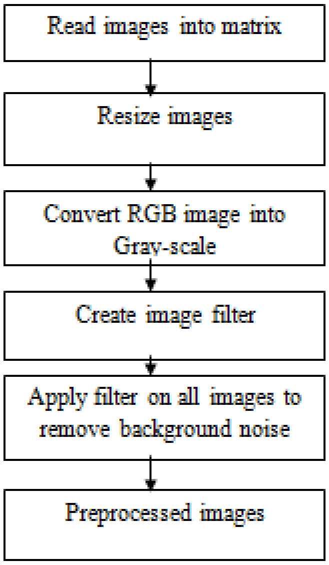
Read images into matrix
Resize images
Convert RGB image into Gray-scale
Create image filter
Apply filter on all images to remove background noise
Preprocessed images

Figure 3.2: The structure of the preprocessing engine

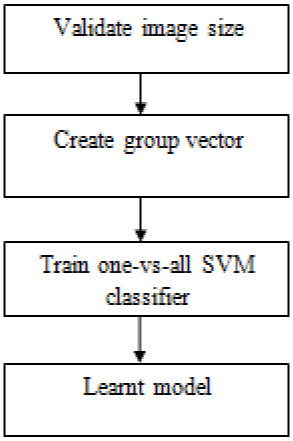

Figure 2.3 The structure of the SVM classifier

ISSN: 2277-9655
Impact Factor: 3.785

Algorithm 1: Indoor Scene Recognition Algorithm

Read the source image, and Extract the features from the source indoor scene image. Feature descriptor will be the sub image, and will describe smaller details than the original Target image.

1. Perform pre-processing step to validate the feature descriptor set and arrange all of the feature descriptors in the single feature sets as the training set.
2. Prepare the group data by adding the group IDs corresponding with all of the samples or feature descriptors in the training set.
3. Run SVM training on the feature descriptor training set and return the weight and bias information for all feature descriptors in the training set.
4. Run SVM classifier by submitting the SVM weight and bias data, group data and the testing feature descriptor vector.
5. Return the matching SVM classification information.
6. Evaluate the SVM classification information and return the decision logic.
7. Features are extracted from training image set and the decision logic is returned.
8. Feature vector of training image is fed to learnt model to assign a class label to given image.

RESULT ANALYSIS

Indoor scene acknowledgment is a testing open issue in abnormal state vision. Most scene acknowledgment models that function admirably for outside scenes perform inadequately in the indoor space.

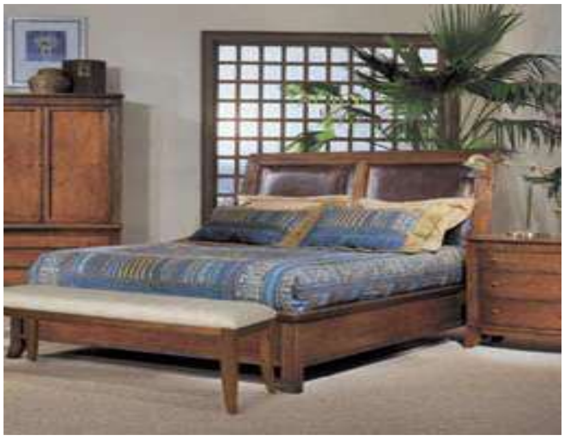

Figure 3.1: The image sample from the indoor scene recognition dataset

The principle trouble is that while some indoor scenes (e.g. passages) can be very much portrayed by worldwide spatial properties, others (e.g., book shops) are better described by the items they contain.

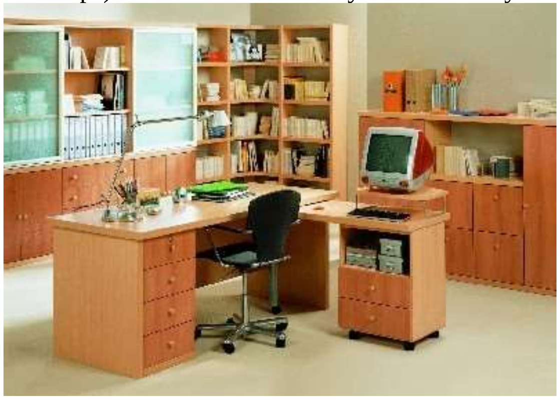

Figure 3.2: Another image from the indoor scene recognition dataset

All the more by and large, to address the indoor scenes acknowledgment issue we require a model that can misuse nearby and worldwide discriminative data.

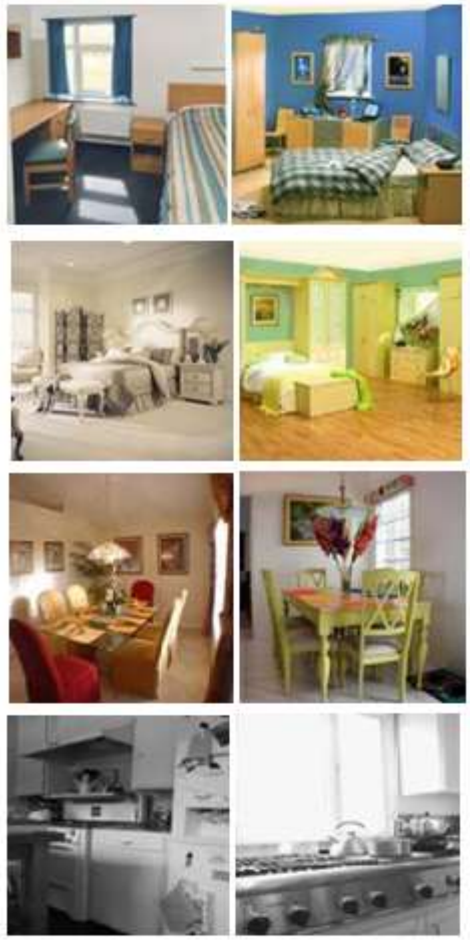

Figure 3.3: The image dataset containing images from different indoor scene categories

In this segment we portray the dataset of indoor scene classes. Most present papers on scene acknowledgment center on a decreased arrangement of indoor and outside classifications. Conversely, our dataset contains countless scene classes. The pictures in the dataset were gathered from different sources: online picture inquiry devices (Google and Altavista), online photograph offering locales (Flickr) and the LabelMe dataset. Fig. 3.3 demonstrates the 67 scene classes utilized as a part of this study. The database contains 15620 pictures. All pictures have a base determination of 200 pixels in the littlest pivot.

This dataset represents a testing characterization issue. As a representation of the in-class variability in the dataset, fig. 3.3 shows normal pictures for some indoor classes. Note that these midpoints have not very many unmistakable traits in examination with normal pictures for the fifteen scene classifications dataset and Caltech 101. These midpoints propose that indoor scene characterization may be a hard undertaking.

The MIT indoor dataset carries the images from total of 67 image categories. The dataset has been shortlisted to the 5 categories with system supportable dataset size for each category. The following table shows the classification performance of our idea using deformable parts-based model (DPM), GIST-color (GC) features and Spatial Pyramid features (SP) on all 67 scene categories. The last column of the table lists the classification performance obtained from associating DPM, GC and SP features (All).

ISSN: 2277-9655
Impact Factor: 3.785

Assuming the positive evaluation set is a trustworthy representation of real world indoor scenes and arranged in the form of different sized images which includes the grayscale, colored or binarized images. The indoor scene dataset has been obtained from MIT and contains multiple images in the four primary categories of bedroom, living room, dining room and office. The 15 test images have been taken from the training set images, whereas 5 images have been taken from image sources other than the training image set. The current 2-stage indoor scene recognition is so sensible with the testing images from the training set and returns the results accuracy at 53.33% approximately. The testing images taken from sources other than training images have returned the 40% correct results. Without any lowering of the threshold this figure has been noticed at some 50% overall accuracy.

PROPERTY NAME	**RESULTS**
Positive	37
Negative	3
Sensitivity	92.59%
Specificity	7.69%
Positive Likelihood Ratio	1.00
Negative Likelihood Ratio	0.96
Prevalence	67.50%
Positive Predictive Value	67.57%
Negative Predictive Value	33.33%

Table 1: The table of properties calculated from the hypothesis of statistical type-1 and type-2 errors on the Category-A testing dataset.

125 images from 5 test subjects were obtained to test the above systems. The data for testing the fully automated indoor scene detection system, manual scene detection and automated scene recognition system and the fully automated scene detection and recognition consists of multiple indoor scenic views of various places in the office, house or corridors. The first image was taken under 'good' conditions with relatively constant lighting conditions with high aperture and low F-value camera for clear image. This would be used as the known as the clear view image in the indoor scene recognition system. The environment condition of the image was categorized by the researcher as 'A'. The other indoor scene database images were taken under worsening conditions with adverse lighting conditions and sometimes with lower aperture and high F-value cameras. These would be used as test images for the indoor scene recognition system. An effort was made to vary the lighting as much a possible in the environment which the images were gathered to test the systems' robustness. The environment condition of the image was categorized as 'B'. Data for the indoor scene recognition was gathered as follows. Nine known images from each individual were collected and three (unknown) images taken when the subject was posing in intermediate angles between the nine known images.

IMAGE CATEGORY	**RECALL (TP/TP+FP)**	**PRECISION (TP/FN+TP)**
Office	**62.5%**	**83.33%**
Gallery	**75%**	**85.71%**
Living Room	**70%**	**100%**
Kitchen	**77.78%**	**87.50%**
Average	**71.32%**	**89.14%**

Table 2: Category wise Precision and Recall calculation

CONCLUSION

Assuming the positive evaluation set is a trustworthy representation of real world indoor scenes and arranged in the form of different sized images which includes the grayscale, colored or binarized images. The indoor scene dataset has been obtained from MIT and contains multiple images in the four primary categories of bedroom, living room, dining room and office. The 45 test images have been taken from the training set images, whereas 5 images have been taken from image sources other than the training image set. The current 2-stage indoor scene recognition is so sensible with the testing images from the training set and returns the results accuracy at 62.5% approximately. The testing images taken from sources other than training images have returned the 60% correct results. Without any lowering of the threshold this figure has been noticed at some 61.25% overall accuracy. The experimental results have proved to be efficient on the basis of obtained results from the proposed model simulation. The proposed model has been proved to be efficient than the proposed model by approximately 10%. The previous model was recorded with the accuracy nearly at 53%, whereas the proposed model has been recorded at approximately 62.5% for the given dataset. Also, the proposed model has been tested with some of the out of the dataset image, where the proposed model has been produced nearly 60% of accuracy.

REFERENCES

[1] Espinace, Pablo, Thomas Kollar, Nicholas Roy, and Alvaro Soto. "Indoor scene recognition by a mobile robot through adaptive object detection." Robotics and Autonomous Systems 61, no. 9 (2013): 932-947.

[2] Giannoulis, Dimitrios, Dan Stowell, EmmanouilBenetos, Mathias Rossignol, Mathieu Lagrange, and Mark D. Plumbley. "A database and challenge for acoustic scene classification and event detection." In Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European, pp. 1-5. IEEE, 2013.

[3] Antanas, Laura, M. Hoffmann, Paolo Frasconi, TinneTuytelaars, and Luc De Raedt. "A relational kernel-based approach to scene classification." InApplications of Computer Vision (WACV), 2013 IEEE Workshop on, pp. 133-139. IEEE, 2013.

[4] Gupta, Saurabh, Pablo Arbelaez, and Jitendra Malik. "Perceptual organization and recognition of indoor scenes from rgb-d images." In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 564-571. IEEE, 2013.

[5] Juneja, Mayank, Andrea Vedaldi, C. V. Jawahar, and Andrew Zisserman. "Blocks that shout: Distinctive parts for scene classification." In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 923-930. IEEE, 2013.

[6] Monadjemi, Amir, B. T. Thomas, and Majid Mirmehdi. Experiments on high resolution images towards outdoor scene classification. Technical report, University of Bristol, Department of Computer Science, 2002.

[7] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. John Wiley & Sons,, 1999.

[8] Fitzpatrick, Paul. "Indoor/outdoor scene classification project." Pattern Recognition and Analysis.

[9] Quattoni, Ariadna, and Antonio Torralba. "Recognizing indoor scenes." In*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 413-420. IEEE, 2009.

[10] Li, Li-Jia, Hao Su, Yongwhan Lim, and Li Fei-Fei. "Objects as attributes for scene classification." In *Trends and Topics in Computer Vision*, pp. 57-69. Springer Berlin Heidelberg, 2012.

[11] Antanas, Laura, Marco Hoffmann, Paolo Frasconi, TinneTuytelaars, and Luc De Raedt. "A relational kernel-based approach to scene classification." In*Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pp. 133-139. IEEE, 2013.

[12] Mesnil, Grégoire, Salah Rifai, Antoine Bordes, Xavier Glorot, YoshuaBengio, and Pascal Vincent. "Unsupervised and Transfer Learning under Uncertainty-From Object Detections to Scene Categorization." In *ICPRAM*, pp. 345-354. 2013.

[13] Zhang, Lei, Xiantong Zhen, and Ling Shao. "Learning object-to-class kernels for scene classification." *Image Processing, IEEE Transactions on* 23, no. 8 (2014): 3241-3253.

[14] Li, Li-Jia, Hao Su, Li Fei-Fei, and Eric P. Xing. "Object bank: A high-level image representation for scene classification & semantic feature sparsification." In *Advances in neural information processing systems*, pp. 1378-1386. 2010.

[15] S. Vasudevan, R. Siegwart, Bayesian space conceptualization and place classification for semantic maps in mobile robotics, Robotics and Autonomous Systems 56 (2008) 522–537.

[16] P. Espinace, T. Kollar, A. Soto, N. Roy, Indoor scene recognition through object detection, in: IEEE International Conference on Robotics and Automation, 2010.

[17] P. Viswanathan, T. Southey, J. Little, A. Mackworth, Automated place classification using object detection, in: Canadian Conference on Computer and Robot Vision, 2010.