Volume 7, Issue 2, February 2017 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Adaptive Indoor Scene Classification with Multi-SVM Classification to solve Multi-Class Problem

Neetu Bala, Harpreet Singh

Patiala Institute of Engineering & Technology, for Women, Patiala,
Punjab, India

DOI: 10.23956/ijarcsse/V7I2/0147

Abstract: In recent years, indoor scene recognition has attracted much attention and its research has rapidly expanded by not only 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 proposed 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 great challenge. Fortunately, the images used in this project have some degree of uniformity thus the detection algorithm can be simpler: first, the all the faces are vertical and have frontal view; second, they are under almost the same illuminate condition. 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

I. 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. 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. Scene classification is highly valuable in remote navigation also.

Indoor scenes are cluttered with many objects. So classification techniques simply based on color, texture and intensity are not very effective to classify indoor scenes. Pioneering works used SIFT, SURF etc in combination with supervised learning. But these techniques fail to distinguish many indoor scenes. One way to bridge semantic gap between image representation and image recognition is to make use of more and more sophisticated models, but good learning and inference is extremely difficult task for such models. Alternatively semantic gap between low-level features like color, intensity, texture etc. and high-level category label can be reduced by introducing object-based representation as intermediate representation. As the performance of scene recognition is heavily dependent on feature representation, this object-based intermediate representation proves to be useful in enhancing classification results. Recently objects-based techniques for indoor scene classification have proven to be showing promising performance over other state-of-art techniques.

In this work, we will review the recent and significant techniques that have been used for indoor scene classification. Besides we will identify the key approaches being used in indoor scene classification. The major contributions made by each significant work and the challenges posed to efficient classification will also be discussed.

II. LITERATURE REVIEW

Espinace, Pablo and Thomas Kollar[1] have worked on Indoor scene recognition by a mobile robot through adaptive object detection. In this paper authors have proposed a new technique to achieve this goal. As a distinguishing feature, authors used common objects, such as Doors or furniture, as a key intermediate representation to recognize indoor scenes. Authors have framed our method as a generative probabilistic hierarchical model, where they have used object category classifiers to associate low level visual features to objects, and contextual relations to associate objects to scenes. The inherent semantic interpretation of common objects allows us to use rich sources of online data to populate the probabilistic terms of our model. In contrast to alternative computer vision based methods, authors boost performance by exploiting the embedded and dynamic nature of a mobile robot. In particular, they have increased detection accuracy and efficiency by using a 3D range sensor that allows us to implement a focus of attention mechanism based on geometric and structural information. Giannoulis, Dimitrios and Dan Stowel[2] have worked on a project based upon database and challenge for acoustic scene classification and event detection. In this paper authors have introduced a newly-launched public evaluation challenge dealing with two closely related tasks of the field: acoustic scene classification and event detection. Authors gave an overview of the tasks involved; describe the processes of creating the dataset; and define the evaluation metrics. Finally, illustrations on results for both tasks using baseline methods applied on this dataset are presented, accompanied by open-source code. Antanas, Laura and M. Hoffmann[3] have developed a relational kernel-based approach to scene classification. In this paper authors have shown that relational techniques can also improve scene classification. More specifically, we employ a new relational language for learning with kernels, called kLog. With this language authors defined higher-order spatial relations among semantic objects. When applied to a particular image, they characterize a particular object arrangement and provide discriminative cues for the scene category. The kernel allows us to tractably learn from such complex features. Thus, our contribution is a principled and interpretable approach to learn from symbolic relations how to classify scenes in a statistical framework. Gupta, Saurabh, Pablo Arbelaez, and Jitendra Malik[4] have proposed perceptual organization and recognition of indoor scenes from rgb-d images. The authors have addressed the problems of contour detection, bottom up grouping and semantic segmentation using RGB-D data. They have focused on the challenging setting of cluttered indoor scenes, and evaluate our approach on the recently introduced NYU-Depth V2 (NYUD2) dataset [27]. They have proposed algorithms for object boundary detection and hierarchical segmentation that generalize the gPb − ucm approach of by making effective use of depth information. They have also shown that our system can label each contour with its type (depth, normal or albedo). We also propose a generic method for long-range amodal completion of surfaces and show its effectiveness in grouping. Juneja, Mayank et. al.[5] have worked on blocks that shout: distinctive parts for scene classification. In this paper, authors have proposed a simple, efficient, and effective method to do so. We address this problem by learning parts incrementally, starting from a single part occurrence with an Exemplar SVM. In this manner, additional part instances are discovered and aligned reliably before being considered as training examples. Authors have also proposed entropy-rank curves as a means of evaluating the distinctiveness of parts shareable between categories and use them to select useful parts out of a set of candidates.

III. EXPERIMENTAL DESIGN

Indoor scene classification system involves various steps like feature extraction, feature selection, feature vector generation, training and classification. The flow chart in figure 4.1 demonstrates all the phases system goes through.

The first phase of the system involves the data preparation phase. The data preparation phase involves the gathering the indoor image dataset from available sources. The actual design of the system starts with gathering data for experimentation. The collected images are then transformed into an appropriate format suitable for system design.

The MIT-indoor dataset has been utilized for the proposed model testing, which is the standard benchmark for scene classification. It contains a total of 15620 images for 67-indoor scenes like bedroom, living room, inside mall, conference hall, bakery etc. All the images are of variable size. The proposed system uses some of these classes of MIT-indoor dataset for training as well as testing. The training and testing images have been saved in matrix. From each selected class equal number of training samples and testing samples has been taken to avoid biasness of classifier.

Preprocessing phase involves following steps:

i. Read image into matrix:- Reading the image into matrix is the very first step involved in image processing. The size of matrix is proportional to size of image. Each cell of matrix stores the corresponding pixel value of image.
ii. Resize the image:- All images from training and testing dataset are normalized to same size.
iii. Convert RGB images to gray-scale- The RGB images are converted into gray scale images to make the model computationally efficient.

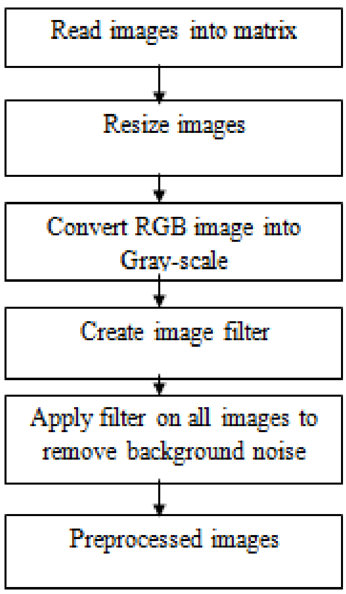


Figure 4.2: preprocessing phase

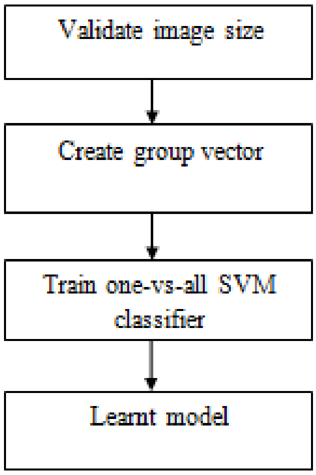
iv. Create image filter:-The Gaussian filter is applied on the image in order to remove the Gaussian noise from the indoor scene test object after converting it into the gray scale..
v. Apply filter:- The filtered images become more robust to feature extraction and classification. The image is convolved with Gaussian filter.
vi. Feature extraction

The following is the feature matching and classification algorithm for matching the extracted indoor scene image with the different images of same scene, which are taken at different times, from different viewpoints, or by different sensors.

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.

IV. 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. 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. 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.

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 distinctive sources: online picture inquiry devices (Google and Altavista), online photograph offering locales (Flickr) and the LabelMe dataset. Fig. 5.3 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. 5.3 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 scene dataset is available online on web.mit.edu/torralba/www/indoor.html. 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 lists the classification performance of our approach using deformable parts-based model (DPM), GIST-color (GC) features and Spatial Pyramid features (SP) on each of the 67 scene categories. The last column of the table lists the classification performance obtained from combining DPM, GC and SP features (All).

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.

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

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%

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.

Table 3: Category wise Precision and Recall calculation

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%**

V. 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, Emmanouil Benetos, 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, Tinne Tuytelaars, 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, Tinne Tuytelaars, 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, Yoshua Bengio, 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.

