INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

*(Volume2, Issue6)*

Available online at: www.ijariit.com

# Multivariate Indoor Scene Recognition using the Object Level Analysis with SVM Classification

**Neetu Dhingra**

*Patiala Institute of Engineering and Technology, India*

***ABSTRACT-- The research area of the indoor scene recognition has attracted the various scientists and engineers across the globe, which includes the neuroscientists, electronics engineers, robotic engineers, digital image experts, camera developers and manufacturers for the purpose of application designing in the fields of the computer vision, vision based communications and the access control systems. The indoor scene recognition methods require the inclusion of the various methods in the computer vision, image processing and feature recognition for the scene recognition by identifying the category of the input image by comparing it against the given training databases by the means of the feature descriptor (popularly based upon the color or low level features) and the classification algorithm. The indoor scene classification algorithms require the number of the computations and feature transformations along with the normalization and automatic categorization. In this thesis, the multi-category dataset has been incorporated with the robust feature descriptor using the scale invariant feature transform (SIFT) along with the multi-category enabled support vector machine (mSVM).***

***KEYWORDS—Indoor scene recognition, object level detection, Scene Aura based Recognition, Image Retrieval.***

## 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, as better as it helps in better organization and browsing of the image data. 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 Stowell[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 k Log. 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.

Monadjemi, Amir et. al.[6] have performed the experiments on high resolution images towards outdoor scene classification. Authors have examined the use of high frequency features in high resolution images to increase texture classification accuracy when used in combination with lower frequency features. They used Gabor features derived from sections of 4032 _ 2688 images. A neural network classifier was used to determine the classification performance of lower and high frequency features when used separately and then in combination. Feature shuffling and Principal Component Analysis was applied to determine both the role of each feature in the classification and to extract a smaller reduced feature set involving both lower and high frequency features.

Duda, Richard O., Peter E. Hart, and David G. Stork [7] have worked on pattern classification for robotic vision based scene classification. The authors have presented the basic results and definitions from linear algebra, probability theory, information theory and computational complexity that serve as the mathematical foundations for the pattern recognition techniques discussed throughout this book.

Fitzpatrick, Paul [8] has developed an Indoor/outdoor scene classification project. Authors have strongly correlated with categories we could imagine treating statistically. He has replicated the human labeling of green and sky, the data suggests we could achieve success rates of around 86%.

 

Quattoni et al. [9] have remarked that actual state of the art algorithms failed on recognizing indoor scenes due to large variability across different exemplars within each class. They have presented an approach to use prototypes to detect regions of interest (ROIs) for recognizing indoor scenes. Each prototype consisted of manually annotated ROIs for each prototype. Small spatial windows were used for searching each ROI in image. Li-Jia Li et al. [10] have jointly worked upon to investigate the contribution of objects to scene classification. This paper discusses the usefulness of image representation based on objects in providing complementary information to the low-level features. In this work a large number of 'object filters' have been used. They have used object-bank of 200 pre-trained SVM object detector at 12 different detection scales and 3 spatial pyramid levels.

L. Antanas et al. [11] have proposed the use of relational language to make the classification results more accurate. They have used the relational language with kernel learning and named it as *KLog*. They have used E/R model for representing relations between objects and spatial relation between objects have been derived from their bounding boxes. The gist of the scene also has been used along with object-bank to capture global properties.

G. Mesnil et al. [12] have focused on reducing curse of dimensionality involved in object-bank. They have proposed to use PCA (principle component analysis) to reduce the dimensionality in combination with deep learning using CAE (Contractive Auto-Encoders). They have proved that the dimensionality can be reduced to a large extent while preserving semantic information of the scene.

L. Zhang et al.[13] have represented each image using object-to-class(O2C) distances, which they have defined as distance of an object to its nearest object in the class. This representation is not only compact but also carries more discriminative feature space.

L. Li et al.[14] have discussed in detail the importance of each component like different detectors, spatial location, scale classication models etc. for efficient scene recognition. They have provided guidelines for using high-level object representation for scene classification.

M. Alberti et al.[15] have presented an approach for joint object detection for indoor scene classification. They have considered directional, distance and size-based relationship between object-pairs. Further, learning is based on Gaussian Mixture Model (GMM).

## III. FINDINGS OF LITERATURE REVIEW

The scene classification is the term used to name a technique which can differentiate the surrounding environment and objects in the vision using some specific or a set of computer vision techniques together. Using the scene classification modules, the computer vision based (camera vision based) devices can automatically detect the type of the scenes. In the previous researches, the researchers have created an adequate number of scene classification algorithms. In one of the existing and effective technique, the authors (P. Espinace et. al.) have developed a scene classification algorithm on the basis of object detection for various applications. This algorithm uses the adaptive objective analysis method in the first stage and takes a final decision about the object types afterwards after complete analysis. The algorithm lacks in labeling the scenes in the different categories. The various applications can be programmed to take better actions if they will know their surroundings better. The existing algorithm is good for door or other rectangular and differently colored objects, but fails at same color object differentiation and the scene type recognition. The correct scene type recognition can help us to design the computer vision based applications with different movement modules. When we need the indoor awareness for computer oriented applications, the indoor scene recognition is essential which must be able to incorporate the multiple object based analytical feature-set, which has been proposed in the proposed model. As in the case of office, it should be capable of moving by seeing and classifying the office furniture, which is very different from the home environment. The office furniture includes the tables and chairs in a large density than a home. If the computer vision dependent application is walking on the road or pedestrian path, it must be aware of the objects out there in order to smoothly run the computer vision applications.

## IV. EXPECTED PROPOSED DESIGN

At first stage, a detailed literature study would be conducted on the methods or architectures about the computer vision and scene classification. In addition, the basic problems and requirement analysis of scene classification models would be thoroughly studied and developed. Literature study will lead us towards refining the structure of the proposed security solution design. Afterwards, the proposed solution will be implemented in MATLAB simulator and a thorough performance analysis would be performed. Obtained results would be analyzed and compared with the existing techniques. In this research, we are proposing a new paradigm in the computer vision or computer vision based scene classification. The proposed algorithm will be a hybrid method of scene classification, which will use object analysis and labeling along with deep background analysis and scene classification on the basis of surrounding type. In our proposed work, we will be using an adaptive object analysis on the basis of its characteristics. These characteristics of the objects will help the computer vision based applications to differentiate between the objects of similar color and similar shape. This will empower the computer vision based applications to detect the objects with more accuracy which will also help it to detect the type of the surroundings, whether it's a home, office or something else. The background classification will give a larger projection about the scene which will help us to categorize the surroundings on the basis of various properties. The background analysis will improve the scene classification and categorization to boost the surrounding specific movement functions in the various vision applications. The performance of the proposed model will be evaluated on the basis of performance parameters like elapsed time, true positive, true negative, false positive, false negative, accuracy, probability of detection, probability of false alarm, etc.

## 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.

[15] Alberti, Marina, John Folkesson, and Patric Jensfelt. "Relational approaches for joint object classification and scene similarity measurement in indoor environments." In *AAAI 2014 Spring Symposia: Qualitative Representations for Robots*. 2014.

[16] Russakovsky, Olga, Yuanqing Lin, Kai Yu, and Li Fei-Fei. "Object-centric spatial pooling for image classification." In *Computer Vision–ECCV 2012*, pp. 1-15. Springer Berlin Heidelberg, 2012.

 

[17] 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.

[18] C. Fredembach, M. Schroder, S. Susstrunk, Eigenregions for image classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (12) (2004) 1645–1649.

[19] A. Mojsilovic, J. Gomes, B. Rogowitz, Isee: perceptual features for image library navigation, in: SPIE Human Vision and Electronic Imaging, 2002.

[20] T. Kollar, N. Roy, Utilizing object–object and object–scene context when planning to find things, in: International Conference on Robotics and Automation, 2009.

[21] P. Dollar, Z. Tu, P. Perona, S. Belongie, Integral channel features, in: British Machine Vision Conference, 2009.

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

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

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

[25] L.-J. Li, H. Su, E. Xing, L. Fei-Fei, and Object bank: a high-level image representation for scene classification and semantic feature sparsification, in: Neural Information Processing Systems, 2010.

[26] S. Thorpe, C. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381 (1996) 520–522.

 