Master's Thesis Pre-Proposal: (Indoor versus Outdoor Scene Classification)

Preethi Rao Vantaram

February 10, 2011

Over the last few years the interest in the research problem of indoor vs. outdoor scene classification[2] has grown significantly, due to its importance in many applications such as Content Based Image Retrieval (CBIR) or Query by Image Content (QBIC), image data organization, robotics, and photography, thus providing a strong motivation for this project. For an example, knowledge of a scene type helps in event classification which is a fundamental component for automatic album systems. The main aim of this project is to implement and analyze two different methodologies for scene classification. More specifically this work is formulated as a binary problem of classifying an arbitrary scene as being indoor or outdoor.

One of the techniques being currently explored is based on edge analysis [3] for indoor/outdoor scene classification is being explored. The hypothesis of this edge based indoor/outdoor classification scheme is that indoor scenes tend to comprise a large portion of objects with straight edges, while the content in outdoor imagery tend to have a significant portion of edges that are curvy. To this effect an edge straightness based approach can be significantly helpful in differentiating between indoor and outdoor scenes. To this effect a straightness metric for an edge is defined, and the percentage of straight edges globally or locally are analyzed to facilitate effective classification of indoor/outdoor scenes. Straightness is measured as the ratio of the length of the straight lines between the first and last points of the edge to the length of the edge in pixels The main advantage of the algorithm in [3] is that is it computationally inexpensive unlike standard classification algorithms such as ones using neural and Bayesian and networks.

On the hand, the Bag of Words (BoW)[4] model is another approach that has been employed effectively for scene classification. The fundamental hypothesis for a classification approach based on the Bag of Words (BoW) methodology is that indoor and outdoor scenes can be effectively identified through uniquely existent features in them, which can be comprehensively represented though a dictionary of their content. Consequently, a BoW approach typically pools together a dictionary of visual words that represents a set of characteristic/basic elements which are distinct for indoor and outdoor images, and can be used to uniquely identify them. This methodology typically commences in a training phase, wherein

visual words that are characteristic to indoor and outdoor scenes are computed to form a dictionary. More specifically, a feature detection algorithm such as the Opponent Scale Invariant Feature Transform (Opponent SIFT) [1] is commonly used to extract features that are invariant to view point, illumination, scale, occlusion, etc, which eventually are utilized to form the visual words. Following this, these features are subjected to a clustering algorithm, and the centroids of the obtained clusters are pooled to form the visual dictionary. In the testing phase of the classification procedure, images to be classified are each represented using the bag of visual words, by the frequency of their occurrence in these images in the form of vectors and subsequently a binary classifier (e.g. SVM) is trained on these vectors.

In this work both the aforementioned approaches for indoor/outdoor scene classification will be scrutinized. It is expected that the BoW method will outperform the edge based approach because of its generic nature of modeling indoor vs. outdoor characteristics. Furthermore, the hypothesis that outdoor scenes have a low percentage of edges that are straight is expected to fail, especially for images pertaining to urban environments (eg. skylines) and large structural content. A confusion matrix will be computed to furnish the classification accuracy for both methods, where in the mean of the classification rates along the main diagonal would provide an indication of the mean classification accuracy across both categories. Metrics such as precision and recall will also be employed if necessary. Since the indoor/outdoor scene classification is a binary classification problem a Receiver Operating Characteristic (ROC) analysis may also be employed to improve the classification accuracy.

The images utilized in this project will be personally collected, as well as from publicly available Torralba and LabelMe databases. The Torralba database contains a total of 15620 jpg format images with 67 Indoor categories. The LabelMe database is more comprehensive collection of imagery comprising of about 10 million indoor/outdoor scenes. For both these databases only a small subset of images will be used in this project.

References

[1] Eugene Mbanya, Christian Hentschel, Sebastian Gerke, Mohan Liu, Andreas Nrnberger, and Patrick Ndjiki-nya. P.: Augmenting bag-of-words - category specific features and concept reasoning. In *In: Working Notes of CLEF 2010*, 2010.

[2] Aude Oliva and Antonio Torralba. Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research*, page 2006, 2006.

[3] Andrew Payne and Sameer Singh. Indoor vs. outdoor scene classification in digital photographs. *Pattern Recogn.*, 38:1533–1545, October 2005.

[4] Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of ACM SIGMM Workshop on Multimedia Information Retrieval*, 2007.