

Indoor vs Outdoor Classification of Consumer Photographs Using Low-Level and Semantic Features

Jiebo Luo

Imaging Science and Technology Lab
Eastman Kodak Company
Rochester, New York 14650

Andreas Savakis

Department of Computer Engineering
Rochester Institute of Technology
Rochester, New York 14623

Abstract

Scene categorization to indoor vs outdoor may be approached by using low-level features for inferring high-level information about the image. Low-level features such as color and texture have been used extensively in image understanding research, however, they cannot solve the problem completely. In this paper, we propose the use of a Bayesian network for integrating knowledge from low-level and mid-level features for indoor vs outdoor classification of images. Using ground truth data for sky and grass detection, we demonstrate that the classification performance can be significantly improved when mid-level features are employed in the classification process.

1. Introduction

Scene categorization is important in a number of applications that deal with consumer photographs. Knowledge of the scene type is useful in event classification, which constitutes a fundamental component of automatic albuming systems [1]. Scene categorization is also valuable in image retrieval from databases because it provides understanding of scene content that can be used along with color, texture, and shape for database browsing. When images are processed through a complex imaging chain, processing operations may be adjusted depending on scene type so that the best rendering can be achieved. The general problem of automatic scene categorization is difficult to solve and is best approached by a divide-and-conquer strategy. A good first step is to consider only two classes such as indoor vs

outdoor [2,3], which may be further subdivided into city vs landscape [4,5], etc.

Scene categorization is often approached by computing low-level features, which are processed with a classifier engine for inferring high-level information about the image [2,4]. In [4], color and texture features were computed for the entire image or for image subsections. One of the issues when dealing with a diverse set of features is how to integrate them into a classification engine. The solution proposed in [4] was to independently classify image subsections and obtain a final result using a majority classifier.

One problem with the methods using low-level features in scene categorization is that it is often difficult to generalize these methods to diverse image data beyond the training set. More importantly, they lack semantic image interpretation that is extremely valuable in determining the scene type. Scene content such as the presence of people, sky, grass, etc., may be used as cues for improving the classification performance obtained by low-level features alone [3]. Sky and grass regions can be identified using color and texture features and classifiers that are tuned for sky and grass detection [3].

In this paper, we propose a Bayesian network approach for the integration of low-level color and texture features and mid-level sky and grass features. This approach improves the classification performance over using low-level features alone. In the following sections, we present an overview of Bayesian networks and propose a system for indoor vs outdoor classification that integrates low-level and mid-level features.

2. Overview of Bayesian Networks

Bayesian or belief networks provide an effective knowledge representation and inference engine in artificial intelligence and can be used in a variety of image understanding applications [6]. Bayes networks are directed, acyclic graphs that encode the cause-effect and conditional independence relationships among variables in the probabilistic reasoning system [7]. The network structure can easily incorporate domain-specific knowledge and a complicated joint probability distribution can be reduced to a set of conditionally independent relationships that are easier to characterize. Thus, a Bayes network can be used to represent the dependence relationships between various features that are represented by random variables at the nodes of the network. The directions of links represent causality and the links between the nodes, or variables, represent the conditional probabilities of inferring the existence of one variable (destination) given the existence of the other variable (source).

Probabilistic reasoning uses the joint probability distribution of a given domain to answer a question about this domain. According to Bayes' rule, the posterior probability can be expressed by the joint probability, which can be further expressed by conditional probability and prior probability:

$$P(S|E) = \frac{P(S,E)}{P(E)} = \frac{P(E|S)P(S)}{P(E)},$$

where S denotes the semantic task and E denotes evidence. With Bayes networks, the computation of the joint probability distribution over the entire system given partial evidences about the state of the system is greatly simplified by using Bayes' rule to exploit the conditional independence relationships among variables. A Bayes net consists of four components: (a) Priors: The initial beliefs about various nodes in the Bayes net; (b) Conditional probability matrices (CPMs): Knowledge about the relationship between two connected nodes in the Bayes net; (c) Evidences: Observations from feature detectors that are input to the Bayes net; and (d) Posteriors: The final computed beliefs after the evidences have been propagated through the Bayes net.

A Bayes network can be viewed as a knowledge representation because it encodes the joint probability distribution. It can also be considered as an inference engine because its evaluation produces the posterior joint probability distribution given evidence of various variables. Bayesian networks offer several advantages: explicit uncertainty characterization, fast and efficient computation, and quick training. They are highly adaptive and easy to build, and provide explicit representation of domain-specific knowledge in human reasoning framework. We found that for our applications, Bayes networks offer good generalization with limited training data, easy maintenance when adding new features or new training data, and convenience in building performance-scalable versions by pruning features.

Training Bayes nets is facilitated by the assumption that each network link is independent of other links at the same level. Therefore, it is convenient to train the entire net by training each link separately, i.e., deriving the CPM for a given link independent of others. Two methods are used for obtaining the CPM for each root-feature node pair: expert knowledge and contingency tables. With the expert knowledge method, an expert is consulted to obtain the conditional probabilities of each feature detector. The contingency tables approach is a sampling and correlation method where multiple observations of each feature detector are recorded and compiled together to create contingency tables which, when normalized, can be used as the CPM. The conditional probability matrix for a link can be trained using frequency counting only when *ground truth* for the parent node is available. Ground truth refers to knowing the correct label of each training sample with some degree of certainty.

3. Indoor vs Outdoor Classification

The Bayesian network structure shown in Figure 1 is proposed for classification of images to indoor vs outdoor. The network integrates low-level features (color and texture) and mid-level features (sky and grass) using a single classification engine. The conditional

probability matrices for each node were derived using the frequency counting approach based on a Kodak database of consumer images [2]. The color features are based on the quantized color histogram (3 x 64 bins) in the Ohta color space [2]. The texture features were based on the Multiresolution Simultaneous Autoregressive (MRSAR) model [2]. The classification based on color or texture was based on the k-nearest neighbor classifier ($k = 1$), and yielded 74% and 82% respectively for a database of 1300 images.

The sky and grass features were extracted using two methods. First the ground truth information about the images was used, i.e. the sky and grass detection is always correct. The indoor/outdoor classification results obtained this way reflect an upper bound, since the performance of any sky and grass classifier will be suboptimal. The second method involves using classification results to obtain sky and grass information. Sky and grass classification methods are based on color/texture features and yield a performance of 95% correct with 10% false positives.

The choice of threshold determines whether the image is indoor or outdoor. When the belief at the root node is above the threshold, the image is characterized as outdoor, thus, the network behaves as an outdoor detector. The threshold value was determined using one-fifth of the available data for training and the remaining data for testing. The value of 0.35 yielded the best overall results for both the training and testing data. In fact, the performance is statistically the same on both data sets.

The indoor/outdoor classification results with ground truth information for sky and grass are shown in Tables 1 and 2, where an overall classification of 90.1% is obtained. The indoor/outdoor classification results when sky and grass detection is based on color/texture classifiers are shown in Tables 3 and 4. In all cases, the use of semantic features improves the system performance and an overall percent correct of 84.7% was obtained when using both low-level and semantic features.

These results provide an improvement over the classification results based on color or texture alone. The major reason for such an improvement is due to the incorporation of mid-

level semantic features. The Bayesian network provides a good framework for integrating all the features, as shown in Tables 2 and 4. It should be noted that the low threshold value indicates that the presence of small evidence of outdoor features (sky and grass) is sufficient reason to classify the image as outdoor.

On-going research involves the development of robust sky and grass classifiers and their use in the existing framework for indoor vs outdoor classification. After the initial classification to indoor or outdoor has been done, further classification to subclasses can be performed, e.g. city, forest, mountain, sea, etc.

4. References

1. A. C. Loui and A. E. Savakis, "Automatic Image Event Segmentation and Quality Screening for Albuming Applications," *Proc. Int. Conf. Multimedia and Expo*, New York, NY, 2000.
2. M. Szummer and R. W. Picard, "Indoor-Outdoor Image Classification", *IEEE International Workshop on Content-Based Access of Image and Video Databases, ICCV '98*, 1998.
3. S. Paek, C. L. Sable, V. Hatzivassiloglou, A. Jaimes, B. H. Schiffman, S.-F. Chang and K. R. McKeown, Integration of Visual and Text Based Approaches for the Content Labeling and Classification of Photographs, *ACM SIGIR '99 Workshop on Multimedia Indexing and Retrieval*, Berkeley, CA, August 19, 1999.
4. A. Vailaya, A. Jain and H. J. Zhang, "On Image Classification: City Images vs Landscapes," *Pattern Recognition*, pp. 1921-1935, 1998.
5. M. Gorkani and R. W. Picard, "Texture Orientation for Sorting Photos at a Glance," *IEEE Conference on Pattern Recognition*, Jerusalem, Israel, Oct. 1994.
6. J. Luo, A. Savakis, S. Etz, and A. Singhal, "On the Application of Bayes Networks to Semantic Understanding of Consumer Photographs," *Proc. ICIP 2000*, Vancouver, Canada.
7. J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, 1988.

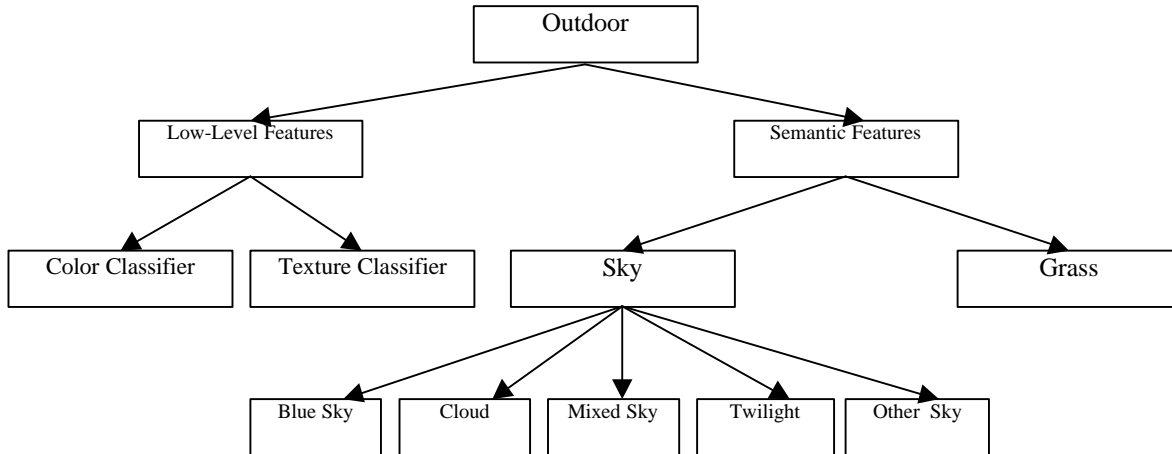


Figure 1: Bayesian Network for Indoor vs Outdoor Classification

Indoor vs Outdoor Classification using “Best Case” Semantic Features			
	Correct	Incorrect	Percent Correct
Indoor	560	55	91.0%
Outdoor	619	74	89.3%
Overall	1179	129	90.1%

Table 1: Indoor vs outdoor classification results with sky and grass classification obtained from ground truth data.

Indoor vs Outdoor Classification using “Best Case” Semantic Features						
	[C]olor	[T]exture	C+T	C+[S]emantic	T+S	C+T+S
Percent Correct	74.2%	82.2%	82.3%	80.9%	86.9%	90.1%

Table 2: Indoor vs outdoor classification results with integration of low-level and mid-level features. Sky/grass results were obtained from ground truth data and color/texture results were based on 1-nn classification using the full image.

Indoor vs Outdoor Classification using Computed Semantic Features			
	Correct	Incorrect	Percent Correct
Indoor	519	96	84.4%
Outdoor	589	104	85.0%
Overall	1108	200	84.7%

Table 3: Indoor vs outdoor classification results with sky and grass classification obtained from ground truth data.

Indoor vs Outdoor Classification using Computed Semantic Features						
	[C]olor	[T]exture	C+T	C+[S]emantic	T+S	C+T+S
Percent Correct	74.2%	82.2%	82.3%	75.2%	84.0%	84.7%

Table 4: Indoor vs outdoor classification results with integration of low-level and mid-level features. Sky/grass results were obtained using color/texture classification and color/texture results were based on 1-nn classification using the full image