Semantic modeling of natural scenes based on contextual Bayesian networks

Huanhuan Cheng*, Runsheng Wang

ATR National Laboratory, Institute of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, China

Abstract

This paper presents a novel approach based on contextual Bayesian networks (CBN) for natural scene modeling and classification. The structure of the CBN is derived based on domain knowledge, and parameters are learned from training images. For test images, the hybrid streams of semantic features of image content and spatial information are piped into the CBN-based inference engine, which is capable of incorporating domain knowledge as well as dealing with a number of input evidences, producing the category labels of the entire image. We demonstrate the promise of this approach for natural scene classification, comparing it with several state-of-art approaches.

1. Introduction

Scene classification, categorizing images into discrete categories (e.g., beach, forest or indoor), is a classical yet challenging problem in computer vision. It is an intermediate step to close the semantic gap between the image understanding of the user and the computer. From the application viewpoint, scene classification is relevant in systems for organization of personal and professional image and video collections. As such, this problem has been widely explored in the context of content-based image retrieval [1], but most prior approaches [2–4] have focused on mapping a set of classic low-level vision features to semantically meaningful categories using a classifier engine.

The semantic modeling of scenes by an intermediate representation was next proposed in order to reduce the gap between low-level and high-level image processing. The meaning of the semantic of the scene is not unique, and two basic strategies based on semantic representation can be found in literatures. One is the object-based strategy [5–8], which identifies the semantic as a set of materials or objects that appear in the image (e.g., sky, grass and rocks). These methods are mainly based on first segmenting the image in order to deal with different regions. Subsequently local classifiers are used to label the regions as belonging to an object. Finally, using this local information, the global scene is classified. Another popular approach is the bag-of-words strategy [9–13], which uses more general intermediate representations. In this case, they first identify a dictionary of visual words or local semantic concepts in order to build the bag-of-words, and further use bag-of-words models (e.g., probabilistic latent semantic analysis (pLSA) [14] and latent Dirichlet allocation (LDA) [15]) to discover clusters of local semantic concepts for scenes.

Although these two strategies have both achieved some promising results, images with similar visual contents are often mis-categorized, particularly for natural scenes. For example, in experiments on Vogel’s dataset of natural scenes [7], coasts and river/lakes are frequently confused, and the reported performance of river/lakes is less than the average rate. This is also observed in the result based on bag-of-words methods [12,16]. Fig. 1 shows two images from Vogel’s dataset. For the first river/lakes scene, the percentages of water, sky and rocks are very similar with those in coasts images. The second image is clearly a forest scene. However, the amount of grass causes the image high probability to be classified as a plain scene. In other words, common materials of scenes usually produce similar semantic features. For this reason, even the approaches based on semantic modeling fail to distinguish them correctly. Therefore, this shows that there is still a challenging work, and more advanced classification methods need to be designed for scene classification.

For natural scenes, a scene is generally composed of several entities, organized in often unpredictable layouts, varying with different seasons and weathers. This makes natural scenes hard to be distinguished. However, without any accurate features extracted from images, people can categorize images into natural scenes very well. Human perception mainly relies on domain knowledge about certain scenes, which includes various attributes such as objects’ occurrence probabilities and their spatial

*Corresponding author. Tel.: +86 731 4575724; fax: +86 731 4518730.
E-mail address: ch2huan@nudt.edu.cn (H. Cheng).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.06.004
The rest of the paper is organized as follows: Section 2 discusses related work. Section 3 describes the general framework we explore. The contextual Bayesian network model is presented in Section 4. Classification results are provided and discussed in Section 5. Section 6 concludes the paper.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات