A Bayesian Network Approach to Multi-feature Based Image Retrieval*

Qianni Zhang and Ebroul Izquierdo

Department of Electronic Engineering, Queen Mary, University of London
London, U.K.
{qianni.zhang, ebroul.izquierdo}@elec.qmul.ac.uk

Abstract. This paper aims at devising a Bayesian Network approach to object centered image retrieval employing non-monotonic inference rules and combining multiple low-level visual primitives as cue for retrieval. The idea is to model a global knowledge network by treating an entire image as a scenario. The overall process is divided into two stages: the initial retrieval stage which is concentrated on finding an optimal multi-feature space stage and doing a simple initial retrieval within this space; and the Bayesian inference stage which uses the initial retrieval information and seeks for a more precise second retrieve.

1 Introduction

The general problem of retrieving, classifying and recognizing patterns in images has been investigated for several decades by the image processing and computer vision research communities. Learning approaches, such as neural networks, kernel machines, statistical and probabilistic classifiers, can be trained to obtain satisfactory results for very specific applications. If the structure of the database and relevant low-level features are known then the constrained pattern recognition problem can be solved with relatively high accuracy. Much of the related work on image classification for indexing and retrieval has focused on the definition of low-level descriptors and the generation of metrics in the descriptor space [1], [2]. These techniques are aimed at defining image signatures using primitives extracted from the content, e.g. pixel patterns and dynamics in images and video or sampling patterns in audio signals. These descriptors are extremely useful in some generic image classification tasks or when classification based on query by example is considered. However, if the aim is to annotate single objects in complex images using semantic words or sentences, there are two questions to be answered in order to solve difficulties that are hampering the progress of research in this direction. Firstly, how to deal with the subjective interpretation of images by different users under different conditions? Secondly, how to link semantically meaningful objects in images with low-level metadata?

The first difficulty, originated from the dependence of perceptual similarity on both user and context, can be tackled by applying rule based semantic reasoning, typically expressed as ontology, to facilitate the recognition of concepts based on experts’ rules and experience. The common solution to the second difficulty considers linking a

---

* The work leading to this paper was partially supported by the European Commission under contracts FP6-001765 aceMedia and FP6-027026 K-Space.
semantic concept, e.g., a keyword, with low-level metadata. In order to achieve this, a machine needs to learn associations between complex combinations of low-level patterns and conceptual objects. Consequently, complex combinations of features building high-dimensional and heterogeneous feature spaces need to be considered. Therefore aim of this paper is two-fold: to devise a Bayesian Network approach to object-centered image retrieval employing non-monotonic inference rules and to combine multiple visual primitives as cue for retrieval. In the literature the Bayesian approach is mainly used in region or object-based image retrieval systems [3], [4], [5], [6], in which the object’s likelihood can be calculated from the conditional probability of feature vectors. These systems use probabilistic reasoning to exploit the relationship between objects and their features. Some other systems employ Bayesian approach in scenario of scene classification e.g., sunset, indoor, outdoor, landscape [7], [8]. However, this kind of classifications has been restricted to mutually exclusive categories, and so is only suitable to images that have only on dominant concept. But in more realistic scenarios in image classification and retrieval, images are complex and usually consist of many semantically meaningful objects. Therefore the relationships between semantically meaningful concepts of the objects cannot be ignored and need to be explored in great degree. In [9], multi-categories are introduced with a simple “sub-class-of” relationship between some of the parent categories with their children categories. While in [10], [11], a statistical analysis of the relationships among concepts is adopted and achieve image retrieving by using metadata (e.g. the RDF triples or semantic web ontology model) that describe and organize the concepts in images.

In this paper, the idea is to model a global knowledge network, thus one entire image is treated as a scenario, and the beliefs is formulated regarding concepts in possibly small regions of the entire image. However in our current work, segmentation is not assumed since segmenting an image into single object is almost as challenging as the semantic gap problem itself. A simple approach is taken to deal with objects in images small image blocks of regular size are considered, which are referred to as “elementary building blocks” in the following of the paper. The overall process is divided into two stages: the initial retrieval stage and Bayesian belief network inference stage, as shown in Fig.1. The initial retrieval stage is concentrated on finding an optimal multi-feature space based on a set of well-selected “representative blocks” for each concept, and then a simple initial retrieval is done within this space, providing essential information for the second stage. In this step the Multi-Objective Optimization (MOO) technique is adopted for estimating the ‘optimal’ multi-feature metric space [12], [13], [14]. The Bayesian inference stage models initial believes on probability distributions of concepts from the initial retrieval information and construct a Bayesian belief network. A more precise second- retrieval is conducted by inferring the presence of objects from the interactions between different concepts on image level.

This paper is organized as follows: Section 2 describes how the multi-feature space is constructed using MOO technique, and how the initial block-level retrieval is done. Section 3 introduces the method of building the Bayesian belief network in this scenario and inferring posterior probabilities out of the network. Experimental results are shown in Section 4 and the paper concludes with Section 5.