Chapter 8

3D Model–Based Semantic Categorization of Still Image 2D Objects

Raluca-Diana Petre
TELECOM SudParis and Alcatel-Lucent Bell Labs, France

Titus Zaharia
TELECOM SudParis and UMR CNRS 8145 MAP5, France

ABSTRACT

Automatic classification and interpretation of objects present in 2D images is a key issue for various computer vision applications. In particular, when considering image/video, indexing, and retrieval applications, automatically labeling in a semantically pertinent manner/huge multimedia databases still remains a challenge. This paper examines the issue of still image object categorization. The objective is to associate semantic labels to the 2D objects present in natural images. The principle of the proposed approach consists of exploiting categorized 3D model repositories to identify unknown 2D objects, based on 2D/3D matching techniques. The authors use 2D/3D shape indexing methods, where 3D models are described through a set of 2D views. Experimental results, carried out on both MPEG-7 and Princeton 3D models databases, show recognition rates of up to 89.2%.

INTRODUCTION

The amount of multimedia content (still image, video, 2D/3D graphics, etc.) available today for the general public is continuously increasing due to the spectacular evolution in digital technologies. Within this context, disposing of powerful search and retrieval methods becomes a key issue for efficient indexing and intelligent access to audio-video material. When large databases are involved, user access to specific material of interest is not possible without efficient search engines and tools. Until recently, retrieval tools were exclusively based on keywords. However, the linguistic barriers represent an important drawback of such approaches. Also, a prior manual anno-

DOI: 10.4018/978-1-4666-2940-0.ch008
3D Model-Based Semantic Categorization of Still Image 2D Objects

tation is required, which is a tedious and highly subjective process.

Within this context, the need of automatic object categorization tools appears as a crucial challenge. The objective is to determine automatically the semantic meaning of an object present in an image or video. A large number of existing methods use prior knowledge in order to accomplish such an objective. Such approaches, so-called machine learning (ML) techniques (Mitchel, 1997; Xue et al., 2009), automatically learn to recognize complex structures based on examples. ML techniques involve two main stages. First, some characteristic features are extracted starting from a set of examples involved in a given training dataset. Then, these features are used in order to recognize new cases. Such methods should be able to generalize the features of a given class while ensuring the accuracy of the recognition process.

However, when a large number of categories involved, the number of recognition criteria (and implicitly the number of exploited features) increases and thus the computational complexity may become intractable (Li, 2006). In addition, in order to allow generalization, a large variety of models should be used in the training set. Notably, we have to take into account that even a given object may present very different appearances due to pose variation. Thus, the training set should include not only a variety of examples but also different instances of the same object, corresponding to different poses.

In this paper we present a new recognition method able to deal with a large variety of objects. Instead of using machine learning techniques, exploit categorized 3D models from existing 3D repositories in the classification process.

The paper is structured as follows. The following chapter presents the 2D/3D shape-based indexing approaches, with basic principle and related work. The proposed 2D/3D indexing methodology, with viewing angle selection strategies and adopted 2D shape descriptors, is then detailed. After introducing the recognition framework, we present and discuss the experimental results obtained. Finally, we conclude the paper and open some perspectives of future work.

SHAPE-BASED 2D/3D INDEXING

Let us first present the general principle of 2D/3D indexing methods.

Basic Principle

The principle of 2D/3D indexing approach is to represent a 3D model, denoted by $M$ as a set of 2D views obtained according to different projection angles. The main hypothesis is that if the models are similar, then they should present similar views. On the contrary, if the two 3D models are different, then there is no correspondence between the sets of views.

A main advantage of the 2D/3D indexing methodology is that it allows different types of comparison. Thus, a 3D model can be compared with other 3D models based on their respective projections. In the same time, a 3D model can be compared in this framework with a 2D object extracted from a 2D image.

The set of views is generated using a set of viewing angles (i.e., positions of the camera in the 3D space). Before generating the set of views, the model is normalized in size and position in order to obtain a canonical representation. First, the object is centered in the origin of the Cartesian system and resized such that it fits the unit sphere. Next, a Principal Component Analysis (PCA) (Schwengerdt et al., 1997) is performed in order to compute the axes of inertia of the 3D model. Finally, the rotation invariance is obtained by rotating the 3D model such that its principal axes coincide with those of the coordinate system (Figure 1).