Video Ontology

Jeongkyu Lee
University of Bridgeport, USA

Pragya Rajauria
University of Bridgeport, USA

Subodh K. Shah
University of Bridgeport, USA

INTRODUCTION

There has been a great deal of interest in the development of ontology to facilitate knowledge sharing and database integration. In general, ontology is a set of terms or vocabularies of interest in a particular information domain, and shows the relationships among them (Doerr, Hunter, & Lagoze, 2003). It includes machine-interpretable definitions of basic concepts in the domain. Ontology is very popular in the fields of natural language processing (NLP) and Web user interface (Web ontology). To take this advantage into multimedia content analysis, several studies have proposed ontology-based schemes (Hollink & Worring, 2005; Spyropoulos, Paliouras, Karkaletsis, Kosmopoulos, Pratikakis, Perantonis, & Gatos, 2005). Modular structure of the ontology methodology is used in a generic analysis scheme to semantically interpret and annotate multimedia content. This methodology consists of domain ontology, core ontology, and multimedia ontology. Domain ontology captures concepts in a particular type of domain, while core ontology is the key building blocks necessary to enable the scalable assimilation of information from diverse sources. Multimedia ontology is used to model multimedia data, such as audio, image, and video. In the multimedia data analysis the meaningful patterns and hidden knowledge are discovered from the database. There are existing tools for managing and searching the discovered patterns and knowledge. However, almost all of the approaches use low-level feature values instead of high-level perceptions, which make a huge gap between machine interpretation and human understanding. For example, if we have to retrieve anomaly from video surveillance systems, low-level feature values cannot represent such semantic meanings. In order to address the problem, the main focus of research has been on the construction and utilization of ontology for specific data domain in various applications. In this chapter, we first survey the state-of-the-art in multimedia ontology, specifically video ontology, and then investigate the methods of automatic generation of video ontology.

BACKGROUND

In general, ontology is a set of terms or vocabularies of interest in a particular information domain, and shows the relationships among them. Sharing common understanding of the structure of information is one of the more common goals in developing ontology. Another reason is that it enables reuse of domain knowledge. The other reasons include making domain assumptions explicit, separating domain knowledge from the operational knowledge, and analyzing domain knowledge. For example, once we have ontology for video surveillance camera in one airport, it can be sharable with the other airport surveillance systems, and provide semantic knowledge on the airport system.

Modular structure of the ontology methodology is used in a generic analysis scheme to semantically interpret and annotate multimedia content. This methodology consists of domain ontology, core ontology, and multimedia ontology. Domain ontology is meant to model the content layer of the multimedia data with respect to specific real world domains (Doerr, Hunter, & Lagoze, 2003). Domain ontology languages allow rich ontology structures. It is very important for the

audiovisual industry to have a methodology for the interoperability of a language to understand the operating ability of the domain ontology. Because domain ontology represents concepts in very specific and often eclectic ways, they are often incompatible. As systems that rely on domain ontology expand, they often need to merge domain ontologies into a more general representation. Core ontology is one of the key building blocks necessary to enable the scalable assimilation of information from diverse sources (Hunter, 2003). A complete and extensible ontology that expresses the basic concepts is essential for well-defined mappings between domain-specific knowledge representation and the subsequent building of a variety of services such as cross domain searching, browsing, data mining, and knowledge extraction. The goal of core ontology is to provide a global and extensible model into which data originating from distinct sources can be mapped and integrated. This canonical form can then provide a single knowledge base for cross-domain tools and services. Jane Hunter (2003) proposes the ABC event-aware model that is able to easily ascertain the intersections, differences, and domain-specific aspects of each ontology or multimedia content. In Figure 1, ABC ontology is a core ontology that can be used to describe, record, and differentiate between the domain ontologies, such as museum, multimedia content description, right management, and biomedical ontologies.

Recent advances in computing technologies have made available vast amounts of digital video content, resulting in a growing research interest in the extraction of semantic multimedia content that can provide a description in a conceptual level. Although significant progress has been made on automatic segmentation of video data and the recognition of low-level features within such content, comparatively little progress has been made on machine-generated semantic descriptions of audiovisual information. To this end, several approaches (Hollink, & Worring, 2005; Spyropoulos et al., 2005) address the problem of building video ontology to enable the inclusion and exchange of video content through a common understanding of the content description and semantic information.

Web technology is used to transform a large existing video ontology embedded in an annotation tool into a commonly accessible format (Geurts, van Ossenbruggen, Hardman, & Davis, 2004). The recombination of existing video materials is then used as an example application, in which video metadata enables the retrieval of multimedia content based on both content descriptions and multimedia concepts. The current semantic Web technology provides a readily applicable set of tools and languages for annotating multimedia data. It intents in creating a universal medium for information exchange by putting documents with computer-process able meaning on the Web. The semantic Web is about having data as well as documents on the Web so that machines can process, transform, assemble, and even act on the data in useful ways.

Multimedia annotation is the process of describing images or videos, and retrieval is the process of finding them (Dasiopoulou, Papastathis, Mezaris, Kompatsiaris, & Strintzis, 2004; Meland, Austvik, & Heggland, 2003). The two major approaches are content-based image/video retrieval and the metadata-based approach.

Figure 1. The Harmony Project's ABC model (Hunter, 2003)