Chapter 6

A Multidimensional Approach for Describing Video Semantics

Uma Srinivasan, CSIRO ICT Centre, Australia

Surya Nepal, CSIRO ICT Centre, Australia

ABSTRACT

In order to manage large collections of video content, we need appropriate video content models that can facilitate interaction with the content. The important issue for video applications is to accommodate different ways in which a video sequence can function semantically. This requires that the content be described at several levels of abstraction. In this chapter we propose a video metamodel called VIMET and describe an approach to modeling video content such that video content descriptions can be developed incrementally, depending on the application and video genre. We further define a data model to represent video objects and their relationships at several levels of abstraction. With the help of an example, we then illustrate the process of developing a specific application model that develops incremental descriptions of video semantics using our proposed video metamodel (VIMET).

INTRODUCTION

With the convergence of Internet and Multimedia technologies, video content holders have new opportunities to provide novel media products and services, by repurposing the content and delivering it over the Internet. In order to support such

applications, we need video content models that allow video sequences to be represented and managed at several levels of semantic abstraction. Modeling video content to support semantic retrieval is a hard task, because *video semantics* means different things to different people. The MPEG-7 Community (ISO/ISE 2001) has spent considerable effort and time in coming to grips with ways to describe video semantics at several levels, in order to support a variety of video applications. The task of developing content models that show the relationships across several levels of video content descriptions has been left to application developers. Our aim in this chapter is to provide a framework that can be used to develop video semantics for specific applications, without limiting the modeling to any one domain, genre or application.

The Webster Dictionary defines the meaning of semantics as "the study of relationships between "signs and symbols" and what they represent." In a way, from the perspective of feature analysis work (MPEG, 2000; Rui, 1999; Gu, 1998; Flickner et al., 1995; Chang et al. 1997; Smith & Chang, 1997), low- level audiovisual features can be considered as a subset, or a part of, visual "signs and symbols" that convey a meaning. In this context, audio and video analysis techniques have provided a way to model video content using some form of constrained semantics, so that video content can be retrieved at some basic level such as shots. In the larger context of video information systems, it is now clear that feature analyses alone are not adequate to support video applications. Consequently, research focus has shifted to analysing videos to identify higher-level semantic content such as objects and events. More recently, video semantic modeling has been influenced by film theory or semiotics (Hampapur, 1999; Colombo et al., 2001; Bryan-Kinns, 2000), where a meaning is conveyed through a relationship of signs and symbols that are manipulated using editing, lighting, camera movements and other cinematic techniques. Whichever theory or technology one chooses to follow, it is clear that we need a video model that allows us to specify relationships between signs and symbols across video sequences at several levels of interpretation (Srinivasan et al., 2001).

The focus of this chapter is to present an approach to modeling video content, such that video semantics can be described incrementally, based on the application and the video genre. For example, while describing a basketball game, we may wish to describe the game at several levels: the colour and texture of players' uniforms, the segments that had the crowd cheering loudly, the goals scored by a player or a team, a specific movement of a player and so on. In order to facilitate such descriptions, we have developed a framework that is generic and not definitive, but still supports the development of application specific semantics.

The next section provides a background survey of some of these approaches used to model and represent the semantics associated with video content. In the third section, we present our Video Metamodel Framework (VIMET) that helps to model video semantics at different levels of abstraction. It allows users to develop and specify their own semantics, while simultaneously exploiting results of video analysis techniques. In the fourth section, we present a data model that implements the VIMET metamodel. In the fifth section we present an example. Finally, the last section provides some conclusions and future directions.

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/multidimensional-approach-describing-video-semantics/25971

Related Content

Wireless Emergency Services
Jun Sun (2005). *Encyclopedia of Multimedia Technology and Networking (pp. 1096-1103).*
www.irma-international.org/chapter/wireless-emergency-services/17373

A Semiotic Model of Multimedia: Theory and Evaluation
Helen C. Purchaseand Daniel Naumann (2001). *Design and Management of Multimedia Information Systems: Opportunities and Challenges (pp. 1-21).*
www.irma-international.org/chapter/semiotic-model-multimedia/8113

A-DisETrac Advanced Analytic Dashboard for Distributed Eye Tracking
Yasasi Abeysinghe, Bhanuka Mahanama, Gavindya Jayawardena, Yasith Jayawardana, Mohan Sunkara, Andrew T. Duchowski, Vikas Ashokand Sampath Jayarathna (2024). *International Journal of Multimedia Data Engineering and Management (pp. 1-20).*
www.irma-international.org/article/a-disetrac-advanced-analytic-dashboard-for-distributed-eye-tracking/341792

Towards Robust Invariant Commutative Watermarking-Encryption Based on Image Histograms
Roland Schmitz, Shujun Li, Christos Grecosand Xinpeng Zhang (2014). *International Journal of Multimedia Data Engineering and Management (pp. 36-52).*
www.irma-international.org/article/towards-robust-invariant-commutative-watermarking-encryption-based-on-image-histograms/120125

Adaptive Hypermedia for Personalised TV
Judith Masthoffand Lyn Pemberton (2005). *Adaptable and Adaptive Hypermedia Systems (pp. 246-263).*
www.irma-international.org/chapter/adaptive-hypermedia-personalised/4188