Chapter VIII

Object-Based Video Analysis and Interpretation

Ying Luo, University of Washington, USA

Jeng-Neng Hwang, University of Washington, USA

Tzong-Der Wu, Chinese Culture University, Taiwan

ABSTRACT

In this chapter, we present a novel scheme for object-based video analysis and interpretation based on automatic video object extraction, video object abstraction, and semantic event modeling. In this scheme, video objects (VOs) are first automatically extracted, followed by a video object abstraction algorithm for identifying key frames to reduce data redundancy and provide reliable feature data for the next stage of the algorithm. Semantic feature modeling is based on a temporal variation of low-level features of video objects. Dynamic Bayesian networks (DBNs) are then used to characterize the spatio-temporal nature of the video objects. The system states in the proposed DBNs directly correspond to the physical concepts. Thus, the decoding of the DBN system states from observable variables is a natural interpretation of the behavior.
of the video objects. Since the video objects are generally considered as the dominant semantic features of video clips, the proposed scheme provides a powerful methodology for content description, which is critical for large scale MPEG-7 applications.

INTRODUCTION

The explosive growth of the amount of multimedia information, such as images, audio and video, calls for highly efficient methods for representing, indexing, retrieving and filtering of multimedia contents. Flexible and extensible representation schemes have been proposed with the advent of the MPEG-7 standard. To integrate MPEG-7 description schemes into multimedia applications, fully automatic algorithms for multimedia analysis and interpretation, i.e., content description, are the essential tools needed for large-scale multimedia applications. The discussion in this chapter will focus on the analysis and interpretation of video sequences. Nevertheless, the framework is readily extended to other multimedia applications.

The best method to demonstrate how a human understands video clips is by an example. When we watch the “Miss America” video sequence, our interpretation is that “a woman is talking.” Two specific aspects of the video sequences are emphasized in the human interpretation: (1) It is the object (i.e., a woman) that is used by humans as the basic unit of understanding. (2) The interpretation is in terms of the semantics (i.e., talking) of the objects. It can be summarized that humans interpret the contents of video clips by the semantics of the objects in the video.

In contrast to the easy grasp of the contents of video clips by a human, fully-fledged interpretation of video sequences by computer is an extremely difficult task. The basic signal processing issue, i.e., the extraction of low-level features including both video and audio features such as motion, color and spectral features, is more or less a solved problem. In order for the computers to interpret the video sequences as a human does, there are two big gaps that need to be filled: one is the comprehensive algorithm that extracts the objects from low-level features. The other is the modeling of object behavior to a semantic level. The approach described in this chapter is trying to make a step toward the solving of these two problems.

The objective of this chapter is to describe an automatic video analysis and interpretation framework based on video object (VO) extraction, object-based abstraction and semantic feature mapping. Background about the definitions of the necessary concepts and a review of the literature are presented. The video object extraction and object-based abstraction are then described. Next, the mapping from object features to high-level semantic concepts by dynamic Bayesian networks is discussed. Finally, features of the proposed system are summarized and possible future work is discussed.

BACKGROUND

To design a complete video analysis and interpretation system we must solve two fundamental problems, i.e., what object-based features shall we count on and what mapping shall we use. For the first problem, we have both video and audio information at hand. In this chapter, we will only use video (visual) features extracted from segmented
A Feature Selection Approach for Network Intrusion Classification: The Bi-Layer Behavioral-Based


[www.igi-global.com/article/a-feature-selection-approach-for-network-intrusion-classification/103958?camid=4v1a](www.igi-global.com/article/a-feature-selection-approach-for-network-intrusion-classification/103958?camid=4v1a)