Chapter IX

Semantic Content-Based Retrieval for Video Documents

Lilac Al-Safadi, Saudi Arabia

Janusz Getta
University of Wollognong, Australia

INTRODUCTION

The advancement of multimedia technologies has enabled electronic processing of information to be recorded in formats that are different from the standard text format. These include image, audio and video formats. The video format is a rich and expressive form of media used in many areas of our everyday life, such as in education, medicine and engineering. The expressiveness of video documents is the main reason for their domination in future information systems. Therefore, effective and efficient access to video information that supports video-based applications has become a critical research area. This has led to the development of, for example, new digitizing and compression tools and technology, video data models and query languages, video data management systems and video analyzers. With applications of a vast amount of stored video data, such as news archives and digital television, video retrieval became, and still is, an active area of research.

Why Content-Based Retrieval?

Current video retrieval systems, such as in libraries and news archives, return a whole video document based on search criteria. Sometimes it is not enough to know whether a video document contains a piece of information; it is also important to return the part of the video that contains the required information. For example, in searching the news archives, a user may be interested in the video clip where the president is conducting a speech about the peace process and not the whole event.

Traditional Way of Content-Based Retrieval

The traditional way in searching for part of a video is a tedious and a time-consuming process. The user starts with searching subjective textual description—related to the subject of the video—for a set of keywords or video titles. The process results in making a reference to the matching video document. Users sequentially view the video document

to locate the required clip. Of course this approach is impractical and unrealistic in applications with a vast amount of video data. A system that can *automatically* retrieve a set of video clips from a large collection of pre-stored video documents is needed.

Current Approaches in Addressing Video Content

Currently, a number of approaches have been followed in determining search criteria for retrieving digital video documents. These approaches are based on:

- Media description, such as type, format and compression techniques.
- Content classification, such as user's level of expertise and program category.
- Subjective description, such as keywords, title and producer.
- Technical description, such as length, recording speed, frame number and time.
- Content description, such as casts and their descriptions, actions and relationships.

End users, most of the time, are interested in the context of a video and refer to a video clip by describing its content and not by the low-level technical details. Therefore, users need a retrieval system that aims at randomly retrieving a set of video clips based on the content description, which is referred to as *video content-based retrieval*.

Semantic Content-Based Retrieval

Today, the database world and computer technologies are becoming more human-oriented (centered on the human nature). In this work, by *human nature* we mean the way humans view a video document, extracts and addresses its content, and builds their mental model to describe the video content in order to comprehend. Humans tend to address a video based on contained meanings or semantics. Through retrieval, humans would like to find information in response to spontaneous worded requests and in a way it meets their perception of video document content. Hence, video retrieval systems' new trend aims at retrieving video clips based on semantic content, which is referred to as semantic video content-based retrieval.

> " … what distinguishes one movie from another is the sequence of the events, the story, but not necessarily the sequence of color histograms or edge maps" (Dimitrova, 1995)

Approaches in Video Content-Based Retrieval

Video documents contain two categories of content: perceptual and semantic.

- *Perceptual* content, sometimes referred to as low-level content, is what is seen and heard represented visually in terms of visual features, such as pixels, colors, texture and shape, aurally in terms of audio features, such as loudness, pitches, brightness and frequencies, and textually in terms of alphabets and symbols.
- *Semantic* content is the *meaning* of what has been seen or heard conveyed by the perceptual content.

Throughout this work, *content* will refer to both the perceptual and the semantic content in a video unless specified otherwise.

Consecutively, two main approaches have been followed in video content-based retrieval:

- Perceptual-based retrieval utilizes processing techniques in matching the video database indexes and queries content represented in terms of visual or aural samples or features.
- Semantic-based retrieval searches the video database for meanings similar to those occurring in the user's query.