Chapter VI

Video Content-Based Retrieval Techniques

Waleed E. Farag, Indiana University of Pennsylvania, USA
Hussein Abdel-Wahab, Old Dominion University, USA

ABSTRACT

The increasing use of multimedia streams nowadays necessitates the development of efficient and effective methodologies and systems for manipulating databases storing these streams. These systems have various areas of application such as video-on-demand and digital libraries. The importance of video content-based retrieval (CBR) systems motivates us to explain their basic components in this chapter and shed light on their underlying working principles. In general, a content-based retrieval system of video data consists of the following four stages: (1) Video Shot Boundary Detection, (2) Key Frames (KFs) selection, (3) features extraction (from selected KFs), and (4) retrieval stage (where similarity matching operations are performed). Each one of the above stages will be reviewed and expounded based on our experience in building a Video Content-based Retrieval (VCR) system that has been fully implemented from scratch in JAVA Language (2002). Moreover, current research directions and outstanding problems will be discussed for each stage in the context of our VCR system.

This chapter appears in the book, Multimedia Systems and Content-Based Image Retrieval, edited by Sagarmay Deb. Copyright © 2004, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
INTRODUCTION

Recently, multimedia applications are undergoing explosive growth due to the monotonic increase in the available processing power and bandwidth. This incurs the generation of large amounts of media data that need to be effectively and efficiently organized and stored. While these applications generate and use vast amounts of multimedia data, the technologies for organizing and searching them are still in their infancy. These data are usually stored in multimedia archives utilizing search engines to enable users to retrieve the required information.

Searching a repository of data is a well-known important task whose effectiveness determines, in general, the success or failure in obtaining the required information. A valuable experience that has been gained by the explosion of the web is that the usefulness of vast repositories of digital information is limited by the effectiveness of the access methods (Brunelli, Mich, & Modena, 1999). In a nutshell, the above statement emphasizes the great importance of providing effective search techniques. For alphanumeric databases many portals (Baldwin, 2000) such as google, yahoo, msn, and excite have become widely accessible via the web. These search engines provide their users a keyword-based search model in order to access the stored information but the inaccurate search results of these search engines is a known drawback.

For multimedia data, describing unstructured information (such as video) using textual terms is not an effective solution because they cannot be uniquely described by a number of statements. That is mainly due to the fact that human opinions vary from one person to another (Ahanger & Little, 1996), so that two persons may describe a single image by totally different statements. Therefore, the highly unstructured nature of multimedia data renders keyword-based search techniques inadequate. Video streams are considered the most complex form of multimedia data because they contain almost all other forms such as images and audio in addition to their inherent temporal dimension. The central role of video data among all other multimedia forms motivated us to focus in this chapter on proposing an effective search paradigm for that particular media.

One promising solution that enables searching multimedia data, in general, and video data in particular is the concept of content-based search and retrieval. The basic idea is to access video data by their contents; for example, using one of the visual content features. Realizing the importance of content-based searching, researchers have started investigating the issue and proposing creative solutions (Chang, 1998). Most of the proposed video indexing and retrieval prototypes have the following two major phases (Flinkner et al., 1995):

- Database population phase, consisting of the following steps:
  - Shot boundary detection. The purpose of this step is to partition a video stream into a set of meaningful and manageable segments (Idris & Panchanathan, 1997), which then serve as the basic units for indexing.
  - Key frames selection. This step attempts to summarize the information in each shot by selecting representative frames that capture the salient characteristics of that shot.
  - Extracting low-level features from key frames. During this step, a number of low-level spatial features (color, texture, etc.) are extracted in order to use them as indices to key frames and hence to shots. Temporal features (e.g., object motion) can be used too.
Related Content

Genetic Adaptation of Level Sets Parameters for Medical Imaging Segmentation
[www.irma-international.org/chapter/genetic-adaptation-level-sets-parameters/39559/](www.irma-international.org/chapter/genetic-adaptation-level-sets-parameters/39559/)

Word-Level Multi-Script Indic Document Image Dataset and Baseline Results on Script Identification

Machine Vision Based Non-Magnetic Object Detection and Removal on Moving Conveyors in Steel Industry through Differential Techniques
[www.irma-international.org/article/machine-vision-based-non-magnetic/74801/](www.irma-international.org/article/machine-vision-based-non-magnetic/74801/)

Direct 3D Information Determination in an Uncalibrated Stereovision System by Using Evolutionary Algorithms
[www.irma-international.org/article/direct-information-determination-uncalibrated-stereovision/55098/](www.irma-international.org/article/direct-information-determination-uncalibrated-stereovision/55098/)
Evaluating an Evolutionary Particle Swarm Optimization for Fast Fuzzy C-Means Clustering on Liver CT Images