## Chapter 3

# Correlation–Based Ranking for Large–Scale Video Concept Retrieval

**Lin Lin**
*University of Miami, USA*

**Mei-Ling Shyu**
*University of Miami, USA*

## ABSTRACT

*Motivated by the growing use of multimedia services and the explosion of multimedia collections, efficient retrieval from large-scale multimedia data has become very important in multimedia content analysis and management. In this paper, a novel ranking algorithm is proposed for video retrieval. First, video content is represented by the global and local features and second, multiple correspondence analysis (MCA) is applied to capture the correlation between video content and semantic concepts. Next, video segments are scored by considering the features with high correlations and the transaction weights converted from correlations. Finally, a user interface is implemented in a video retrieval system that allows the user to enter his/her interested concept, searches videos based on the target concept, ranks the retrieved video segments using the proposed ranking algorithm, and then displays the top-ranked video segments to the user. Experimental results on 30 concepts from the TRECVID high-level feature extraction task have demonstrated that the presented video retrieval system assisted by the proposed ranking algorithm is able to retrieve more video segments belonging to the target concepts and to display more relevant results to the users.*

## INTRODUCTION

Multimedia retrieval has become a popular research area due to the explosive growth of digital image and video collections and the widespread accessibility of media in social networks and internet. The demand for solutions and tools to search and retrieve the interesting information effectively and efficiently is increasing. Meanwhile, the capacity of multimedia data grows larger and faster. For instance, it has become more suitable to measure the sizes of videos in TB (terabytes) rather than in GB (gigabytes) now. Hence, how to manage and retrieve the desired information

from the huge amounts of multimedia data has challenged researchers in the multimedia area (Chen, 2010).

Concept-based retrieval (Snoek & Worring, 2008) is to detect the existence of objects (such as bus and hand), the meaning of scenes (such as cityscape and nighttime), and the occurrence of events (such as airplane flying and people dancing). It enables the users to utilize multimedia data for entertainment, distant education, commerce and business, social communication, navigation, security, surveillance, and etc. For example, a user may enjoy watching the segments of videos with singing if she/he loves music, or may seek news videos with protest content if she/he is interested in politics. Correctly detecting the classroom setting from the videos would help information search for educational applications, and retrieving the bridge and mountain would assist the users who are planning a trip. The high-level concepts such as doorway and street from video games could be used for navigation, while emergency vehicle and traffic intersection from video surveillance and security cameras could be used for tracking.

Most of the existing search and retrieval approaches are restricted to textual information which is metadata such as surrounding text and closed caption, or are dependent on an interactive framework which requires users' feedback and log files. The advances of database and data warehouse technologies provide us a proper way to manage these textual data and they seem to be efficient tools that are able to facilitate the users to access the data on demand. However, challenges arise when heavy human efforts are demanded for annotation, correcting the textual information, as well as performance evaluation of the retrieved results. To address these issues, content-based multimedia retrieval has emerged in recent years. Most of the content-based frameworks utilize support vector machine (SVM) detectors trained on scale-invariant feature transform (SIFT) descriptors and rank the retrieved results based on the scores obtained from the classifiers. However,

SVM is very time consuming with a huge demand in space. Moreover, the classification-based ranking methods suffer from the ad-hoc mechanism to determine the threshold for class labels. Therefore, they cannot be used for real-time online searching.

In addition to efficiency, another important consideration of a retrieval system is effectiveness. The overall retrieval performance is usually evaluated through the mean average precision (MAP) of the retrieved results obtained from the ranking algorithm. To make a fair comparison on the effectiveness of the approaches, the benchmarked video concepts provided by the TREC Video Retrieval Evaluation (TRECVID) community (Smeaton, Over, & Kraaij, 2006) are the most commonly used testbed for evaluating large-scale standardizing data sets. In 2008 and 2009, there are totally 30 concepts for high-level feature extraction task and 219 videos with annotations for the training purpose (Divakaran, 2009).

In this paper, a novel correlation-based ranking algorithm is proposed to assist in video concept retrieval from large-scale video collections. There are three stages in our video retrieval system. First, video content is represented by the global and local low-level audio and visual features rather than the point SIFT descriptors, and thus both the global representation of scenes and local representation of regions are captured. Second, multiple correspondence analysis (MCA) is applied to explore the correlation between video content and semantic concepts rather than between video context and semantic concepts, thus without the effort for extracting the textual information from the videos. In our previous work, MCA-based correlation information has been investigated and used as transaction weights for data pruning (Lin, Shyu, & Chen, 2009) and feature weights for generating weighted association rules (Lin & Shyu, 2010). Moreover, MCA technique has been used for low-level feature selection in event detection (Shirahama, Sugihara, Matsuoka, & Uehara, 2009) and concept detection (Zhu, Lin, Shyu, & Chen, 2010). In this paper, MCA has been utilized

## Related Content

Mobile Ad Hoc Networks: Protocol Design and Implementation

Crescenzio Gallo, Michele Perilliand Michelangelo De Bonis (2011). *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts (pp. 31-47).*

www.irma-international.org/chapter/mobile-hoc-networks/50578

Applying Smart Security Model to Protect Client Services From the Threats of the Optical Network

Kamel H. Rahoumaand Ayman A. Ali (2021). *Advancements in Security and Privacy Initiatives for Multimedia Images (pp. 80-113).*

www.irma-international.org/chapter/applying-smart-security-model-to-protect-client-services-from-the-threats-of-the-optical-network/262069

Enhancing Rating Prediction by Discovering and Incorporating Hidden User Associations and Behaviors

Ligaj Pradhan (2019). *International Journal of Multimedia Data Engineering and Management (pp. 40-59).*

www.irma-international.org/article/enhancing-rating-prediction-by-discovering-and-incorporating-hidden-user-associations-and-behaviors/232181

Implement Multichannel Fractional Sample Rate Convertor using Genetic Algorithm

Vivek Jainand Navneet Agrawal (2017). *International Journal of Multimedia Data Engineering and Management (pp. 10-21).*

www.irma-international.org/article/implement-multichannel-fractional-sample-rate-convertor-using-genetic-algorithm/178930

Video Segmentation and Structuring for Indexing Applications

Ruxandra Tapuand Titus Zaharia (2011). *International Journal of Multimedia Data Engineering and Management (pp. 38-58).*

www.irma-international.org/article/video-segmentation-structuring-indexing-applications/61311