Perception-level Evaluation of Video Similarity in Content-based Retrieval Systems

Waleed Farag, Ph.D.
Department of Computer Science, Indiana University of Pennsylvania

Abstract:
The increasing use of multimedia streams nowadays necessitates the development of efficient and effective methodologies for manipulating databases storing them. Moreover, content-based access to multimedia databases requires in its retrieval stage to effectively assess the similarity of video data. This work proposes a new technique for measuring video data similarity that attempts to model some of the factors that reflect human perception notion in evaluating video data similarity. This model presents one step towards designing intelligent content-based video retrieval systems capable of measuring the similarity among video clips in a way similar to what humans do. The performance of the proposed model was tested in terms of recall and precision of the retrieved results where the system yielded very satisfactory values of recall and precision under various testing scenarios.

Introduction

The basic objective of any automated video indexing system is to provide the user with easy-to-use and effective mechanisms to access the required information. For that reason, the success of a content-based video access system is mainly measured by the effectiveness of its retrieval phase. The general query model adapted by almost all multimedia retrieval systems is the QBE (Query By Example). In this model, the user submits a query in the form of an image or a video clip (in case of a video retrieval system) and asks the system to retrieve similar data. QBE is considered to be a promising technique since it provides the user with an intuitive way of query presentation. In addition, the form of expressing a query condition is close to that of the data to be evaluated.

Upon the reception of the submitted query, the retrieval stage analyzes it to extract a set of features then performs the task of similarity matching. In the latter task, the query-extracted features are compared to the features stored into the metadata then matches are sorted and displayed back to the user based on how close a hit is to the input query. A central issue here is how the similarity matching operations are performed and based on what criteria. This central theme has a crucial impact on the effectiveness and applicability of the retrieval system.

In this paper, we discuss the design of the retrieval stage of our VCR (Video Content-based Retrieval) system and shed the light on one of its distinguishing characteristics. The system attempts to make its comparison decisions based on modeling the way humans perform similarity matching of video data. This is achieved by using a number of factors reflecting how humans perceive media similarity; thus, the proposed model overcomes the shortcomings of other approaches. The VCR system has, in addition to the retrieval stage discussed in this paper, three other modules namely, shot boundary detection, key frames selection, and the indexing module. After expounding the proposed similarity matching model including video shot similarity definition and the adapted similarity measuring factors, this article concludes by a thorough performance evaluation with results that support the effectiveness of the developed model.