

Assessing Digital Video Data Similarity

Waleed E. Farag

Zagazig University, Egypt

INTRODUCTION

Multimedia applications are rapidly spread at an ever-increasing rate, introducing a number of challenging problems at the hands of the research community. The most significant and influential problem among them is the effective access to stored data. In spite of the popularity of the keyword-based search technique in alphanumeric databases, it is inadequate for use with multimedia data due to their unstructured nature. On the other hand, a number of content-based access techniques have been developed in the context of image and video indexing and retrieval (Deb, 2004). The basic idea of content-based retrieval is to access multimedia data by their contents, for example, using one of the visual content features.

Most of the proposed video-indexing and -retrieval prototypes have two major phases: the database-population and retrieval phases. In the former one, the video stream is partitioned into its constituent shots in a process known as shot-boundary detection (Farag & Abdel-Wahab, 2001, 2002b). This step is followed by a process of selecting representative frames to summarize video shots (Farag & Abdel-Wahab, 2002a). Then, a number of low-level features (color, texture, object motion, etc.) are extracted in order to use them as indices to shots. The database-population phase is performed as an off-line activity and it outputs a set of metadata with each element representing one of the clips in the video archive. In the retrieval phase, a query is presented to the system that in turns performs similarity-matching operations and returns similar data back to the user.

The basic objective of an automated video-retrieval system (described above) 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; Yoshitaka & Ichikawa, 1999). In this model,

the user submits a query in the form of an image or a video clip (in the 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 with 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 the assessment of video data similarity. Appropriately answering the following questions has a crucial impact on the effectiveness and applicability of the retrieval system. How are the similarity-matching operations performed and on what criteria are they based? Do the employed similarity-matching models reflect the human perception of multimedia similarity? The main focus of this article is to shed the light on possible answers to the above questions.

BACKGROUND

An important lesson that has been learned through the last two decades from the increasing popularity of the Internet can be stated as follows: “[T]he usefulness of vast repositories of digital information is limited by the effectiveness of the access methods” (Brunelli, Mich, & Modena, 1999). The same analogy applies to video archives; thus, many researchers are starting to be aware of the significance of providing effective tools for accessing video databases. Moreover, some of them are proposing various techniques to improve the quality, effectiveness, and robustness of the retrieval system. In the following, a quick review of these techniques is introduced with emphasis on various approaches for evaluating video data similarity.

One important aspect of multimedia-retrieval systems is the browsing capability, and in this context some researchers proposed the integration between the human and the computer to improve the performance of the retrieval stage. In Luo and Eleftheriadis (1999), a system is proposed that allows the user to define video objects on multiple frames and the system to interpolate the video object contours in every frame. Another video-browsing system is presented in Uchihashi, Foote, Girgensohn, and Boreczky (1999), where comic-book-style summaries are used to provide fast overviews of the video content. One other prototype retrieval system that supports 3D (three-dimensional) images, videos, and music retrieval is presented in Kosugi et al. (2001). In that system each type of query has its own processing module; for instance, image retrieval is processed using a component called ImageCompass.

Due to the importance of determining video similarity, a number of researchers have proposed various approaches to perform this task and a quick review follows.

In the context of image-retrieval systems, some researchers considered local geometric constraint into account and calculated the similarity between two images using the number of corresponding points (Lew, 2001). Others formulated the similarity between images as a graph-matching problem and used a graph-matching algorithm to calculate such similarity (Lew). In Oria, Ozsu, Lin, and Iglinski (2001) images are represented using a combination of color distribution (histogram) and salient objects (region of interest). Similarity between images is evaluated using a weighted Euclidean distance function, while complex query formulation was allowed using a modified version of SQL (structured query language) denoted as MOQL (multimedia object query language). Berretti, Bimbo, and Pala (2000) proposed a system that uses perceptual distance to measure the shape-feature similarity of images while providing efficient index structure.

One technique was proposed in Cheung and Zakhor (2000) that uses the metadata derived from clip links and the visual content of the clip to measure video similarity. At first, an abstract form of each video clip is calculated using a random set of images, then the closest frame in each video to a particular image in that set is found. The set of these closest frames is

considered as a signature for that video clip. An extension to this work is introduced in Cheung and Zakhor (2001). In that article, the authors stated the need for a robust clustering algorithm to offset the errors produced by random sampling of the signature set. The clustering algorithm they proposed is based upon the graph theory. Another clustering algorithm was proposed in Liu, Zhuang, and Pan (1999) to dynamically distinguish whether two shots are similar or not based on the current situation of shot similarity.

A different retrieval approach uses time-alignment constraints to measure the similarity and dissimilarity of temporal documents. In Yamuna and Candan (2000), multimedia documents are viewed as a collection of objects linked to each other through various structures including temporal, spatial, and interaction structures. The similarity model in that work uses a highly structured class of linear constraints that is based on instant-based point formalism.

In Tan, Kulkarni, and Ramadge (1999), a framework is proposed to measure the video similarity. It employs different comparison resolutions for different phases of video search and uses color histograms to calculate frames similarity. Using this method, the evaluation of video similarity becomes equivalent to finding the path with the minimum cost in a lattice. In order to consider the temporal dimension of video streams without losing sight of the visual content, Adjero, Lee, and King (1999) considered the problem of video-stream matching as a pattern-matching problem. They devised the use of the vstring (video string) distance to measure video data similarity.

A powerful concept to improve searching multimedia databases is called relevance feedback (Wu, Zhuang, & Pan, 2000; Zhou & Huang, 2002). In this technique, the user associates a score to each of the returned hits, and these scores are used to direct the following search phase and improve its results. In Zhou and Huang, the authors defined relevance feedback as a biased classification problem in which there is an unknown number of classes but the user is only interested in one class. They used linear and nonlinear bias-discriminant analysis, which is a supervised learning scheme to solve the classification problem at hand. Brunelli and Mich (2000) introduced an approach that tunes search strategies and comparison metrics to user behavior in order to improve the effectiveness of relevance feedback.

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/assessing-digital-video-data-similarity/17224

Related Content

Making Enterprise Recorded Meetings Easy to Discover and Share

Shimei Pan, Mercan Topkara, Jeff Boston, Steve Woodand Jennifer Lai (2015). *International Journal of Multimedia Data Engineering and Management* (pp. 19-36).

www.irma-international.org/article/making-enterprise-recorded-meetings-easy-to-discover-and-share/130337

FaceTimeMap: Multi-Level Bitmap Index for Temporal Querying of Faces in Videos

Buddha Shrestha, Haeyong Chungand Ramazan S. Aygün (2019). *International Journal of Multimedia Data Engineering and Management* (pp. 37-59).

www.irma-international.org/article/facetimemap/233863

Storytelling and Narrative Marketing in the Era of Social Media

Sonia Ferrari (2016). *Experimental Multimedia Systems for Interactivity and Strategic Innovation* (pp. 1-15).

www.irma-international.org/chapter/storytelling-and-narrative-marketing-in-the-era-of-social-media/135121

From Single to Multiplayer Mobile Bluetooth Gaming

Daniel C. Doolan, Kevin Duggan, Sabin Tabircaand Laurence T. Yang (2009). *Handbook of Research on Mobile Multimedia, Second Edition* (pp. 584-594).

www.irma-international.org/chapter/single-multiplayer-mobile-bluetooth-gaming/21030

Authorship Detection and Encoding for eBay Images

Liping Zhou, Wei-Bang Chenand Chengcui Zhang (2011). *International Journal of Multimedia Data Engineering and Management* (pp. 22-37).

www.irma-international.org/article/authorship-detection-encoding-ebay-images/52773