VALIDATION EXPERIMENTS ON STRUCTURAL, CONCEPTUAL, COLLECTION, AND ACCESS DESCRIPTION SCHEMES FOR MPEG-7

Ana B. Benitez, Shih-Fu Chang

Department of Electrical Engineering, Columbia University, New York, NY, USA
{ana, sfchang} @ ee.columbia.edu

We have recently contributed to the development and validation of several description schemes for multimedia content, which have been integrated into the emerging MPEG-7 standard. This paper includes summaries of the validation experiments of these description schemes.

INTRODUCTION

MPEG-7 aims to standardize a framework to describe multimedia content to facilitate a broad range of multimedia applications. Among the tasks is the definition of content features - Descriptors (Ds) - and structures relating Ds and DSs - Description Schemes (DSs). The first committee draft of MPEG-7 is scheduled to be completed in October of 2000.

We responded to the call for MPEG-7 DS technology with proposals for representing images, videos, home media, and multimedia collections [1]. Components of these proposals have evolved and been integrated into the MPEG-7 DS eXperimental Model (XM) [3] and Working Draft (WD) [4].

MPEG-7 Ds and DSs are structured on the basis of their functionality into content description, management, organization, and navigation/access. Content description components describe the perceivable content: spatio-temporal structure and conceptual (or semantic) notions. Content management components relate to the management of the content. Content organization tools address content classification, collection description, and modeling. Finally, navigation/access tools describe summaries and variations of content for efficient and personalized browsing and access to content.

In this paper, we present our validation experiments on (1) Segment DSs and Annotation DS, (2) Collection Structure DS, and (3) Media Transcoding Hint DS, which respectively address content structure and concepts, collections, and access. In the following sections, we describe input data, procedure, and implementation for each experiment.

STRUCTURE AND CONCEPT

We have validated the Video Segment, Moving Region, Segment Decomposition, Segment Relation Graph, and Annotation DSs in a video retrieval and visualization application. These DSs describe the spatio-temporal structure of visual content except for the Annotation DS, which assigns semantic annotations. The Video Segment DS represents a set of video frames (e.g. video shot); the Moving Region DS, a spatio-temporal area in a video (e.g. object “car”). The Segment Decomposition DS represents the decomposition of video segments or moving regions into others whereas the Segment Relationship Graph DS describes more general relationships among them (e.g. “before” and “to the left”).

In the experiment, for selected video sequences, moving regions corresponding to semantic objects were segmented and tracked using AMOS [6] (e.g. “soccer player”). AMOS was also used to decompose the objects into homogeneous moving regions, and to extract visual features and spatio-temporal relationships among moving regions. Panoramic views of the video segments were generated using [5]. Finally, moving regions and video segments were manually annotated. All this data was encoded using the DSs under validation.

We implemented a retrieval and visualization application that used the generated descriptions to support (1) query by example based on any combination of visual and semantic features for video segments and moving regions (e.g. retrieve similar videos to this one based on color and texture); (2) query by keyword based on semantic annotations (e.g. retrieve videos with an “elephant”); and (3) advanced visualization of the retrieved results based on panoramic views and segmented objects.

The experiment has shown the above DSs’ generality and completeness to encode descriptions of video data generated by AMOS (which precedes MPEG-7) and flexibility to support advanced retrieval and visualization systems.

COLLECTION

We have validated the Collection Structure DS in a browsing and searching scenario. The Collection Structure DS describes a collection of multimedia documents in terms of Collection Cluster DSs – element groupings- and Cluster Relationship DSs - relationships among groupings, in particular, the Cluster Decomposition DS. Collection clusters are described by statistical and semantic features, among others.

Selected image content was classified using two different classification schemes. The classes in each scheme were ad hoc and generated by MPEG-7 members. Examples of classes were “Strange hair” and “Happiness”. The entire collection was represented as a collection cluster for which two cluster decompositions were specified, one for each classification. Each class was also described as a collection cluster and assigned a semantic annotation corresponding to the name of the class and the mean color histogram for the elements in the class. Then, a description instantiating the Collection Structure DS was generated.

The description of the collection was fed into a browsing and searching system, which allowed users to (1) browse the images, the classes, the classes to which an image belongs, and the images assigned to a specific class; and (2) to retrieve similar classes by example in terms on mean color histogram.

We have demonstrated the usefulness of the Collection Structure DS in describing collections for browsing and for searching applications. The are still numerous components of the Collection Structure DS under specification and/or validation. Examples of proposed attributes are statistics (e.g. distribution functions) and author-generated information (e.g. indexers). Examples of proposed relationships are lexical relationships (e.g. “similar/opposite” and “is a”) and statistical relationships (e.g. “more random”).

ACCESS

We have validated the Media Transcoding Hint DS, in particular, the Utility Function D and Scaling Profile D, using an application that generates and delivers different variations of a video program based on network conditions and user preferences. The utility function indicates multimedia users’ quality or utility (e.g. PSNR) as a function of the occupied resources (e.g. bandwidth). The scaling profile lists the preferred sequence of scaling operations for target resource values (e.g. frame dropping and dynamic rate shaping).

In the experiment, utility functions and scaling profiles were generated for selected video sequences using [2] and encoded with the proposed Media Transcoding Hint DS. The delivery of the video was done by simulating variable network conditions in a testbed composed of a video sever, a dummy net, and an HTML client.

The experiment has shown the usefulness of the Media Transcoding Hint DS for delivering and generating different alternatives of video data to fit network conditions or user preferences. The list of utilities, resources, and scaling operations should be extended.

SUMMARY

We have presented our validation work on structural, conceptual, collection, and access DSs for the MPEG-7 standard. More detailed reports, experiment demos, and other resources are available at the ADVENT’s MPEG-7 web site at http://www.ee.columbia.edu/advent/, Research.

ACKNOWLEDGMENTS

The authors wish to thank S. Paek, D. Zhong, A. Jaimes, P. Bocheck (Columbia University), C.-S. Li, J. R. Smith, L. D. Bergman (IBM), Q. Huang, A. Puri (AT&T Labs), C. Judice (Eastern Kodak), C. Saraceno, B. J. Brown (Starlab), C. Jörgensen (Univ. at Buffalo, SUNY), and other participants in the experiments for their valuable contributions to this work.

REFERENCES

[1] A. B. Benitez, S. Paek, S.-F. Chang, Q. Huang, A. Puri, C.-S. Li, J. R. Smith, L. D. Bergman, C. Judice, "Object-Based Multimedia Description Schemes and Applications for MPEG-7", To appear in Image Communications Journal, Special Issue on MPEG-7, 2000.

[2] R. Liao, P. Bocheck, A. Campbell, and S.-F. Chang, "Content-aware Network Adaptation for MPEG-4", Proceedings of NOSSDAV'99, New Jersey, USA, June 1999

[3] MDS Group, “MPEG-7 Multimedia Description Schemes XM”, ISO/IEC JTC1/SC29/WG11 MPEG99/N3112, Maui, Hawaii, USA, Dec.1999.

[4] MDS Group, “MPEG-7 Multimedia Description Schemes WD”, ISO/IEC JTC1/SC29/WG11 MPEG99/N3113, Maui, Hawaii, USA, Dec.1999.

[5] C. Saraceno, B. J. Brown, "Panoramic Images for Video Indexing", TR199909-01, Starlab, Brussels, Belgium, Sep. 1999.

[6] D. Zhong, S.-F. Chang, "AMOS: An Active System for MPEG-4 Video Object Segmentation", 1998 International Conference on Image Processing, Chicago, Illinois, USA, Oct. 4-7, 1998.