Automatic, Context-of-Capture Based, Categorization, Structure Detection and Segmentation of News Telecasts

Arne Jacobs, George T. Ioannidis (Task leader)
Center for Computing Technologies (TZI), University of Bremen, Germany
{jarne, george.ioannidis}@tzi.de

Martha Larson
Fraunhofer Institute for Media Communication, Sankt Augustin, Germany
matha.larson@imk.fraunhofer.de

Stavros Christodoulakis, Nektarios Moumoutzis
Laboratory of Distributed Multimedia Information Systems and Applications
Technical University of Crete (TUC/MUSIC), Chania, Greece
{stavros, nektar}@ced.tuc.gr

Keywords: Information search and retrieval, modeling multimedia data, ontology design, speech recognition, structural models, video analysis

Research problem

The context of an element within a time series can be conceptualized as a state. In the case of digital libraries, the time series of interest are multimedia streams. If a multimedia stream is a recording of a discussion, the context of any single utterance within can be understood to be the state of the discussion at the point the utterance was spoken. This state can be understood as encoding the subject being discussed and the propositions that have been agreed upon by the parties participating in the discussion. For the purpose of multimedia analysis, it is useful to model a context state with an abstract set of hierarchies. In such a model, which we call a Context-of-Capture (CoC), a particular context is characterized by a set of interrelated concepts described in an ontology. It is possible to infer a CoC using the set of words that have occurred in a discussion, or in any general discourse. Knowing the CoC associated with each captured segment of a discourse makes possible discourse interpretation on a semantic level, previously unattained. In the work presented here we focus our research efforts on a particular type of discourse, namely, news broadcasts.

Automatic audiovisual content segmentation is performed mainly at the syntactic level in several systems today, but only few systems exist that take into account the semantics of the audiovisual content. Furthermore, the CoC concept, which represents the context regarding the information captured in an audiovisual segment (e.g. persons, places, events etc.), is either ignored or superficially utilized.

The CoC may be of great importance, especially for programs like news and telecasts, consisting of totally independent video segments and many topic changes. In these cases, the sudden changes in the CoC denote the end of the current context segment and the beginning of a new one.

In addition, the CoC allows for the automatic assignment of the audiovisual segments detected into appropriate thematic categories, as the CoC of a segment contains adequate information for thematic category determination. Equally important with the recognition of a specific context for segmentation and indexing purposes is also the possibility to associate all the knowledge in the knowledge base that is associated with a context.

The above discussion makes clear the need for generic models for describing CoC and scenarios of context appearance, and their use in recognition, segmentation and structuring of the knowledge bases so that complex queries can be answered.

Task objectives

The objective of this task is to develop a demonstrator for automatic categorization, structure detection and segmentation of news telecasts that utilizes advanced structural models. Segment boundary detection will be assisted by a powerful CoC model, which will be used by the appropriate context detection and context change evaluation mechanisms. The segmentation/structural metadata will be finally exported in MPEG-7 format. A query API and user interface will be provided in order to evaluate the results.

Expected results

In particular, task activities will address the following issues:

Definition of the Context-of-Capture (CoC) model. A powerful model for the CoC and of CoC scenarios will be developed. A CoC will be represented as an ontology based on MPEG7 and OWL interoperability will be pursued . CoC scenarios will be represented as graphs representing structures of possible transitions among CoC during a news telecast. Furthermore, the model will support algorithms for utilizing CoC for identification and knowledge management including its use for recognition and inference. The CoC model will be used to classify news stories into relevant contexts based on their content as specified by the recognition process (see below). Moreover, the identified context may be further used to guide the recognition process and resolve ambiguities that may be present after the first recognition phase.

Development of CoC recognition mechanisms. Appropriate algorithms will be developed for CoC recognition, utilizing image, speech and video text processing for audiovisual feature extraction. Simple audiovisual cues (characteristic color, texture, loudness), extraction of text inserts and higher-level visual features (e.g. faces, indoor/outdoor) will be taken into account. In addition, the spoken audio track will be processed using a speech recognizer that identifies the occurrences of keywords, using syllable-based methods . The speech recognizer is trained to recognize exactly those keywords whose presence makes it possible to enrich the CoC models. We assume that the presence of individual keywords will allow us to infer the relevance of higher semantic concepts for particular context states.

Development of mechanisms for CoC–based segmentation and categorization of telecasts. A syntactic segmentation of a telecast can be obtained using algorithms for both shot detection on the visual signal and speaker and speech/music recognition on the audio signal . In addition, a model of the visual syntax of a broadcast may be used to categorize shots into higher-level syntactic segments (like, e.g., report, credits, or presentation). This segmentation can then be further refined by changes in the CoC, denoting the segment boundaries. This refinement mainly concerns the merging of adjacent syntactic segments that have very similar CoC.

Development of a query API and user interface for evaluation. A query API and user interface will be provided in order to evaluate the results. The query API will be used to search a set of classified news stories and the user interface will provide a user-friendly way to browse the results and evaluate the efficiency of the followed approach.

References

[1] Tsinaraki C., Polydoros P., Christodoulakis S.: “Interoperability support for Ontology-based Video Retrieval Applications”, In the Proceedings of CIVR 2004, Dublin/Ireland, July 2004

[2] Tsinaraki C., Polydoros P., Moumoutzis N., Christodoulakis S.: “Coupling OWL with MPEG-7 and TV-Anytime for Domain-specific Multimedia Information Integration and Retrieval”, In the Proceedings of RIAO 2004, April 2004, Avignon/France

[3] A. Miene, Th. Hermes, G.T. Ioannidis: “Extracting Textual Inserts from Digital Videos”, in Proc. of the Sixth International Conference on Document Analysis and Recognition (IDCAR'01), pp. 1079-1083, Seattle, Washington, USA, IEEE Computer Society, September 10-13, 2001

[4] A. Miene, Th. Hermes, G.T. Ioannidis, R. Fathi, and O. Herzog: “Automatic Shot Boundary Detection and Classification of Indoor and Outdoor Scenes”, In Voorhees, E. M. and Buckland, L. P. (eds.), Information Technology: The 11th Text Retrieval Conference, *TREC 2002*, NIST, pp. 615-620, 2003

[5] S. Osang: “Entwicklung eines Schlüsselworterkennungssystems zur Medienbeobachtung”, Diploma thesis, Fachhochschule Bonn-Rhein-Sieg and Fraunhofer IMK

[6] M. Larson, S. Eickeler: “Using Syllable-based Indexing Features and Language Models to Improve German Spoken Document Retrieval”, In *Proceedings of Eurospeech*, 8th European Conference on Speech Communication and Technology, Geneva, September 2003

[7] K. Biatov, J. Köhler: “An Audio Stream Classification and Optimal Segmentation for Multimedia Applications”. In *ACM Multimedia*, San Francisco, November 2003

[8] J. Löffler, J. Köhler, H. Blohmer, K.-U. Kaup: “Archiving of Radio Broadcast Data Using Automatic Metadata Generation Methods within MediaFabric Framework”, In *Proceedings of the 116th AES Convention*, Berlin, May 2004

[9] L. Bankert, A. Jacobs, A. Miene, Th. Hermes, G.T. Ioannidis, O. Herzog: “An Environment for Modelling Telecast Structures”. In T. Catarci, S. Christodoulakis, A. Del Bimbo (eds.), *AVIVDiLib'05 Proceedings*, pp. 176--179, May 2005.