

Towards a Large Scale Concept Ontology for Broadcast Video

Alexander G. Hauptmann

School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
alex+@cs.cmu.edu

Abstract. Earlier this year, a major effort was initiated to study the theoretical and empirical aspects of the automatic detection of semantic concepts in broadcast video, complementing ongoing research in video analysis, the TRECVID video analysis evaluations by the National Institute of Standards (NIST) in the U.S., and MPEG-7 standardization. The video analysis community has long struggled to bridge the gap from successful, low-level feature analysis (color histograms, texture, shape) to semantic content description of video. One approach is to utilize a set of intermediate textual descriptors that can be reliably applied to visual scenes (e.g. outdoors, faces, animals). If we can define a rich enough set of such intermediate descriptors in the form of large lexicons and taxonomic classification schemes, then robust and general-purpose semantic content annotation and retrieval will be enabled through these descriptors.

Our efforts are substantially broad, as our subject matter is broadcast video, which is almost much unrestricted in terms of content, but includes audio and spoken dialog. In addition, broadcast video has an added layer of 'editing': shots and scenes which are carefully chosen to make the point in a broadcast, but do not directly reflect a reality like surveillance video. We are exploring to what extent broadcast video is amenable to a structured characterization of content (using a large but well-defined lexicon). By necessity, the lexicon will have to be general and broadly applicable, since it will be impossible to give in-depth characterizations of shots for the broad content matter that we are dealing with.

But what are the appropriate lexicon items that would allow a sufficiently rich and general description of the video content in broadcast news, which in effect would constitute a general-purpose ontology for describing video content? Our first challenge is to find a large, broad set of descriptors which will be useful in a large variety of broadcast news content. Together with librarians, video archive specialists and experts in multimedia analysis and knowledge representation, we are attempting the definition of an ontology and creating a vocabulary of about a thousand lexical terms that describe the content of broadcast video. Once we have such semantic descriptors, the next question is whether they are actually useful. We will explore the sufficiency and generality of this set of descriptors for annotation and retrieval of video content. One aspect of this work is to empirically determine the feasibility of automatically identifying these descriptions in appropriate video content. We will also annotate

larger amounts of video to see if the set of derived descriptors is appropriate over a wide range of content, and to provide a reference truth for an annotated video library.

Finding 1000 concepts represented in broadcast news video that can be detected and evaluated necessitates careful lexicon design. The concepts in the lexicon should be useful from a perspective of visual information exploitation. Simultaneously the lexicon must be feasible from the perspective of automatic and semi-automatic detection. The design of the lexicon thus needs to bring together members of the library sciences community, knowledge representation as well as researchers from the multimedia analysis community. The confluence of statistical and non-statistical media analysis with ontologies, classification schemas and lexicons helps place the scalable multimedia semantic concept detection problem in the proper context. Context-sensitive concept detection can also help enhance the detection performance and help the scalability. In designing and evaluating a large scale lexicon the following challenges need to be tackled:

- Interpretation of user needs, finding out what do users want from video archives of broadcast news.
- Rigorous experiments to understand how user needs can be mapped into the components of the lexicon
- Study of automatic concept detection system performance and their impact on retrieval performance and classification of concepts on the basis of performance and relevance.
- Understanding of algorithmic approaches for large scale concept detection.
- Empirical and theoretical study of the impact of detection performance trade-off on the ultimate usability of the lexicon, specifically evaluating detection accuracy vs. retrieval performance.

A year-long workshop is under way, developing recommendations and general criteria for the design of large-scale lexicons for audio-visual content classification in support of systems for searching, filtering, and mining of broadcast video. Our approach is to start with a large, fixed collection of data and explore different types of annotation and lexical labeling for retrieval and description. The resulting lexicon and ontology, if successful, will provide a basis for generations of broadcast news video retrieval and annotation work.