



International Conference on Multimedia & Expo (ICME) 2003
Special Session on "Video Segmentation for Semantic Annotation and Transcoding"

Call For Papers

Semantic annotation provides a suitable way to describe, organize and index stored video and semantic transcoding is one of the key processes to adapt the multimedia content to the user requirements. Video segmentation, object identification and tracking a-priori knowledge of the context played in the video are needed to perform both transcoding and annotation. Multimedia data analysis and fusion provide additional confidence for events and highlight detections.

In order to discuss this field, a special session on *Video Segmentation for Semantic Annotation and Transcoding* will be held as a part **2003 IEEE International Conference on Multimedia & Expo (ICME)**

Topics of the special session are (but are not limited to):

- Semantic video segmentation and annotation
- Object-based coding and transcoding
- Subjective performance evaluation metrics of transcoding policies
- Multimedia analysis and data fusion for effective retrieval

Papers accepted for presentation will be published in a special issue of *Multimedia Tools and Applications Journal*, Kluwer Academics Publishers

Important Dates

Deadline for submission	January 15, 2003
Notification of paper acceptance/rejection	March 1, 2003
Final paper submission	March 31, 2003

Paper submission procedure

Papers should be submitted electronically through the Paper Submission System of the main conference (<http://www.icme2003.com/papers/submission.asp>). When uploading a paper, the author should click on the corresponding special session that will appear along with topical areas of regular papers. Note that both initial and final papers are subjected to the same page limit as regular papers, i.e. four-page manuscript in double-column format including authors' names, affiliations, and a short abstract. You can find the complete Paper Kit at the address <http://www.icme2003.org/Papers.asp>.

Session Organizers

Rita Cucchiara, Università di Modena e Reggio Emilia - Italy
(email: cucchiara.rita@unimore.it, phone: 0039-059-2056136 – fax: 0039-059-2056129)
Alberto Del Bimbo, Università di Firenze – Italy
(email: delbimbo@dsi.unifi.it, phone: 0039-055-4796262 – fax: 0039-055-4796363)

Session Secretary

Andrea Prati, Università di Modena e Reggio Emilia - Italy
(email: prati.andrea@unimore.it, phone: 0039-059-2056142 – fax: 0039-059-2056129)

List of invited papers

- | | |
|-----------------|---|
| Authors | A. Vetro, T. Haga, K. Sumi, H. Sun
<i>Mitsubishi Electric Research Laboratories - USA</i> |
| Title | Object-Based Coding for Long-Term Archive of Surveillance Video |
| Abstract | This paper describes video coding and segmentation techniques that can be used to achieve significant increase in storage capacity. Specifically, we examine the possibility to use object-based coding for efficient long-term archiving of surveillance video. We consider surveillance systems with many camera sources in which we are required to store several months of video data for each source, thus storage capacity is a major concern. The paper considers several automatic segmentation algorithms. With each algorithm, we will analyze the shape coding overhead and implication on overall storage requirements, as well as the effect each algorithm has on the reconstructed quality of frames. Additionally, this paper presents techniques to dynamically control the temporal rate of objects in the scene and addresses bit allocation issues. Experimental results show that up to 90% savings in storage can be achieved with the proposed method compared to frame-based video coding techniques. The cost for this savings is that the accuracy of the background is compromised; however, we feel that this is satisfactory for the application under consideration. |
|
 | |
| Authors | N. Peyrard, P. Bouthemy
<i>IRISA / INRIA - France</i> |
| Title | Motion-based selection of relevant video segments for video summarization |
| Abstract | We present a method for motion-based video segmentation and segment classification as a step to video summarization. The aim is to provide a concise and meaningful overview of the video by selecting relevant video segments according to their motion content. The method is composed of three stages. First, a sequential segmentation of the video is performed, corresponding to the detection of changes in the dominant image motion (which can be assumed to be related to the camera motion, i.e., static camera, zooming, panning, tracking,...). It is achieved by analysing the temporal variation of some coefficients of the global 2D affine motion model (robustly) estimated between every two successive images. The video segments thus obtained supply reasonable temporal regions within which to apply a recognition and selection algorithm. The second stage is performed off-line and consists in learning the pre-identified classes of dynamic events (depending on the genre of the processed videos) from a given training set of video samples. We deal now with the real motion content of the scene depicted by the video and no more with the dominant camera motion. We adopt a statistical representation of the motion content, relying on the distribution of temporal co-occurrences of local motion-related measurements. Finally, each video segment is labeled with one of the learned classes of dynamic events (including an extra class for non significative motion content) according to a Maximum a Posteriori (MAP) criterion. Excerpts of the relevant classes are then selected for video summarization. |
|
 | |
| Authors | W. H.-M. Hsu, Shih-Fu Chang
<i>Department of Electrical Engineering - Columbia University - USA</i> |
| Title | A Statistical Framework for Fusing Mid-Level Perceptual Features in News Story Segmentation |
| Abstract | News story segmentation is essential for video indexing, summarization, and navigation in personal video navigators and intelligent media management systems. Most existing systems rely on heuristic rule based methods, in which special cues during story transition were used. In this paper, we will present a new system using a statistical model fusing mid-level “perceptual” features (without relying on the story boundary markers in the closed captions), such as overlay caption continuity, speaker shots, acoustic consistency, significant phrases detected from speech, etc. The model utilizes a family of weighted, exponential functions of binary features to estimate the story boundary likelihood of each hypothesis point. The fusion parameters are estimated by using Kullbak-Leibler divergence measure computed from empirical news video corpora. In order to select the subset of most salient features, an incremental feature inducing procedure is used. The proposed statistical framework for selecting and fusing features in story segmentation can also be used to characterize and distinguish different news production styles in different countries and channels. We will report the performance of such statistic approach against a base-line system, in which anchorperson shots were detected by constrained face detection/clustering and used to detect story boundaries. |