Video Collection Summarization by Semantic Graph Comparison

Philippe Ercolessi, Christine Sénac, Hervé Bredin and Philippe Joly
SAMoVA, IRIT, CNRS UMR 5505, UPS, Toulouse, France
{*ercoless, senac, bredin, joly*}*@irit.fr*

As the amount of available digital video content is increasing exponentially, novel ways of storing, accessing and retrieving it are being developped, such as indexing, segmentation or abstraction techniques. Video abstraction can be useful in many ways – from automatic home movie editing to easier (and faster) exploration of a video collection. Video summaries can be a set of carefully selected key frames or the concatenation of video skims [1].

Rather than focusing on a single video, we aim at summarizing several videos from a single collection. A video collection can be seen as a homogeneous set of videos, with the same type of structure, style, duration, etc. Given a video collection, the objective is to generate a short video that is a representative of this collection. The representativeness of a video depends on the application (faster browsing or narrative summary, for instance). It is not reasonable to process each element of a collection independently to produce the final collection summary. As a matter of fact, such an approach would not take redundancy between videos into account and there is no way it can uncover the commonalities between episodes. Therefore, our approach considers a collection as a whole and rely on two steps: video structuring and structure comparison.

1 Video structuring

1.1 Multimodal segmentation

Videos are segmented based on various visual and audio descriptors. For instance, videos can be divided into *logical story units* [1] based on color histograms. Motion descriptors can be used to split videos based on action scenes, while speaker diarization is performed using MFCC acoustic descriptors. It is also possible to look for various ambiences (music, speech, noises, etc.) [3] or recurring characters [2].

1.2 Representative graph

Based on these segmentations, a clustering step is applied on segments of the whole collection. Semantically similar segments are gathered in the same cluster and temporal relationships between them lead to the definition of one graph per video and type of segmentation.

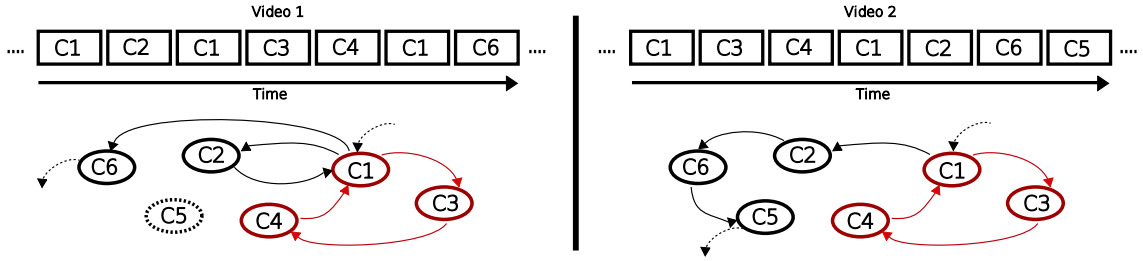

Figure 1: Building and comparing two *semantic* graphs. After the initial segmentation step, each video segment is associated to the closest cluster C_i. *Semantic* graphs are then built by simply linking temporally consecutive segments. Red clusters and links constitute a sub-graph that is common to both videos. Cluster C_5 is absent in first video.

2 Structure comparison

The comparison step consists in finding similarities between the graphs generated during the first step, in order to detect what could be considered as a common characteristic of the whole collection, or to the contrary, what is video-specific. Figure 1 shows an example of graph generation and comparison. This automatic process has to run without any a priori knowledge on the nature of the collection. Therefore, methods as generic as possible are implemented.

3 Collection abstraction

Once common characteristics or specific parts of a collection are clearly determined, it is then possible to generate a video summary which will focus on reccuring parts of the collection, or on specific parts of each video (in the case one summary per video is needed).

References

[1] Sergio Benini, Pierangelo Migliorati, and Riccardo Leonardi. Statistical Skimming of Feature Films. In *International Journal of Digital Multimedia Broadcasting, Volume 2020*. Hindawi Publishing Corporation, 2009.

[2] Elie El Khoury, Christine Sénac, and Philippe Joly. Face-and-clothing based People Clustering in Video Content. In *MIR '10: Proceedings of the International Conference on Multimedia Information Retrieval*, pages 295–304, New York, NY, USA, 2010. ACM.

[3] Julien Pinquier and Régine André-Obrecht. Audio Indexing: Primary Components Retrieval - Robust Classification in Audio Documents. *Multimedia Tools and Applications*, 30(3):313–330, septembre 2006. ISSN 1380-7501 (Print) 1573-7721 (Online).