

# VIDEO SEGMENTATION AND INDEXING

*Moncef Gabbouj*

Institute of Signal Processing  
Tampere University of Technology  
P.O. Box 553, FIN-33101 Tampere, Finland  
Email: moncef.gabbouj@tut.fi

## ABSTRACT

“A lot has been achieved in video segmentation and indexing, but we still have a long way to go.” Video segmentation has been the subject of study for several decades; while, video indexing has mostly been active in the past ten years. Each of these areas has its own motivation and applications, but the fact remains that the success of video indexing depends a great deal on good quality video segmentation. Or is this only a myth?

## 1. INTRODUCTION

The past two decades have witnessed a large growth in research on video segmentation. In addition to individual and group work elsewhere, there have been several research consortia in Europe active in this field. Among them, one can cite EU COST 211 [3] and MoMuSys [10]. On the other hand, work on video indexing started when content-based indexing and retrieval began to emerge in the early 1990s. This paper highlights the major challenges and cites some milestones achieved.

## 2. VIDEO SEGMENTATION

Video segmentation is in general a difficult and complex problem. In certain special conditions, e.g. blue screen or computer generated video, object segmentation can be achieved easily. In any case and assuming that segmentation is successful, semantic (real object) segmentation can still present major challenges.

Video segmentation can be divided into temporal segmentation and spatio-temporal segmentation [7]. Both are used in coding and analysis. The former seeks boundaries between sets of frames where the video contents exhibits some changes; whereas the latter segments and tracks object across frames. All video coding standards include some type of temporal segmentation in the encoder. Some simple techniques use

frame histogram differences. Other ones can be more complicated.

COST 211 contributed to automatic video segmentation during two of its phases, 211 ter and 211 quat. The main contribution was the COST Analysis Model, known as COST AM [1] [3] [4]. This is a rule-based system whose input is a video sequence and its output is an “object” mask and the background. One can obtain additional information from the output of the COST AM, but the segmentation mask is the main result. Supervised or semi-automatic video segmentation is achieved e.g. by the user marking a set of objects of interest and the system proceeds with the segmentation. The user may also stop the segmentation process and “adjusts” some boundary of an object manually. COST AM work continues in project Qimear [11] where the aim is to build a modular framework for video segmentation and analysis.

Due to the lack of ground truth, it is often difficult to optimize or train a segmentation algorithm due to the fact that it is difficult to feedback segmentation errors in order to improve performance. There have been a number of studies to objectively measure segmentation performance but their usefulness remains at the comparative level, i.e. they tell us which technique performs better than another one, but they do not tell us how to improve it. Further challenges are presented by partially occluded objects. Even if these are segmented correctly, their use in indexing and retrieval remains limited, see [13] for possible solutions.

## 3. VIDEO INDEXING

MPEG-7 [8] [9] has given a great push forward for content-based indexing and retrieval work. As in the case of content-based image indexing, video indexing includes low-level indexing and semantic indexing. Additionally, video indexing includes video summarization, “story” segmentation and eventually audio indexing.

Low-level features are usually extracted from video “key-frames.” These can be e.g. color, texture, shape, or spatial

layout. One can of course extract similar features from a group of frames, if needed, but processing is usually performed frame-wise. Semantic indexing, on the other hand, tries to extract features of “meaningful” objects in the video, such as faces, people or other specific objects. Semantic indexing relies heavily on video segmentation before feature extraction. Actually, its success is very tied to good or suitable segmentation. Due to the challenges in video segmentation mentioned earlier, semantic video indexing remains an open problem.

Some success has been achieved in video summarization and story segmentation, but more remains to be done. TREC and TREC VID [12] are among the leading groups in this field.

The audio track is an essential component of a video and it is often “neglected” in video indexing papers. However, audio may contain important information which can be used for video indexing or assist video indexing. The main reason behind the lack of interest in audio indexing is perhaps due to the low number of research papers dealing with indexing of compressed audio. More and more papers are now published dealing with feature extraction in compressed audio, such as MP3 or AAC, [5].

Since video can come at different bitrates, it is of interest to find out at which bitrate video indexing can be performed without significant reduction in retrieval performance, see [6].

#### 4. CONCLUSIONS

Good video segmentation may assist in achieving reliable video indexing. In addition to the challenges associated with video segmentation, the audio track, when available, must be used to assist video indexing. Last, but not least, video indexing schemes must be tested at different bitrates to find out which “features” are to be extracted reliably at which bitrates.

#### 5. REFERENCES

[1] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, T. Sikora, “Image Sequence Analysis for Emerging Interactive Multimedia Services - The European COST 211 Framework,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 7, November 1998, pp. 802-813.

[2] Faouzi Alaya Cheikh, Bodgan Cramariuc, Mari Partio, Pasi Reijonen, and Moncef Gabbouj, “Ordinal-Measure Based Shape Correspondence”, *Journal on Applied Signal Processing*, Vol. 2002, No. 4, April 2002.

[3] COST 211 website: [www.iva.cs.tut.fi](http://www.iva.cs.tut.fi) .

[4] Moncef Gabbouj, Geoff Morrison, Faouzi Alaya-Cheikh, Ronald Mech, “Redundancy Reduction Techniques and Content Analysis for Multimedia Services - the European COST 211 Action,” *Proc. Workshop on Image Analysis for Multimedia Interactive Services 1999 (WIAMIS '99)*, Berlin, Germany, May/June 1999.

[5] Moncef Gabbouj, Serkan Kiranyaz, Kerem Caglar, Esin Guldogan, Olcay Guldogan and Farooq Ahmad Qureshi, “Audio-based Multimedia Indexing and Retrieval Scheme in MUVIS Framework,” *Proceedings of 2003 IEEE International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2003*, (invited plenary talk), Awaji Island, Japan, 7-10 December 2003

[6] E. Guldogan and Olcay Guldogan, "Compression Effects on Content-based Multimedia Indexing and Retrieval Using Color and Texture Attributes", *M.Sc. Thesis*, Tampere University of Technology, Tampere, Finland, January 2003.

[7] I. Kompatsiaris and M. G. Strintzis, “Spatiotemporal Segmentation and Tracking of Objects for Visualization of Videoconference Image Sequences,” *IEEE Trans. on Circuits and System for Video Technology*, vol. 10, no. 8, December 2000.

[8] *Introduction to MPEG-7: Multimedia Content Description Language*, B.S. Manjunath, Philippe Salembier, Thomas Sikora, Eds., Wiley, 2003.

[9] Jose Martinez, “ISO/IEC JTC1/SC29/WG11 N4674: MPEG-7 Overview,” approved document.

[10] MoMuSys home page: <http://www.tnt.uni-hannover.de/project/eu/momusys/>

[11] Qimera homepage, <http://www.qimera.org/>

[12] TREC VID homepage: <http://www-nlpir.nist.gov/>

[13] M. Trimeche, F. Alaya Cheikh and M. Gabbouj, “Similarity Retrieval of Occluded Shapes Using Wavelet-Based Shape Feature,” *SPIE International Symposium on Internet Multimedia Management Systems (VV10)*, Boston, Massachusetts, USA, November 5-8, 2000.