A Collaborative Approach to Video Summarization

Emilie DUMONT[1], Bernard MERIALDO[1], Slim ESSID[2], Werner BAILER[3], Daragh BYRNE[4], Hervé BREDIN[4], Noel E. O'CONNOR[4], Gareth J.F. JONES[4], Martin HALLER[5], Andreas KRUTZ[5], Thomas SIKORA[5], Tomas PIATRIK[6]

[1] EURECOM, Sophia-Antipolis, FRANCE
[2] TELECOM ParisTech, Paris, FRANCE
[3] JOANNEUM RESEARCH, Graz, AUSTRIA
[4] Dublin City University, Dublin, IRELAND
[5] Technische Universität Berlin, Berlin, GERMANY
[6] Queen Mary University of London, UK
{dumont, merialdo}@eurecom.fr, slim.essid@telecom-paristech.fr, werner.bailer@joanneum.at, oconnorn@eeng.dcu.ie, sikora@nue.tu.berlin.de, tomas.piatrik@elec.qmul.ac.uk

Abstract. This poster describes an approach to video summarization based on the combination of several decision mechanisms provided by the partners of the K-Space European Network of Excellence. The system has been applied to the TRECVID 2008 BBC rushes summarization task.

Keywords: Video summarization, segmentation, redundancy removal, selection, fusion

1 Introduction

Video Summarization is a process that requires multimedia analysis of the video sequence to identify redundancies and select the most important and unique parts of the video. Several partners of the K-Space European Network of Excellence [1] have teamed to construct a summarization system based on a combination of feature analyses. This system has been applied to the TrecVid BBC Rushes Summarization task. The approach that has been defined is based on three steps:

- First, a common segmentation of the video is produced by fusing several segmentations based on various indicators, and including confidence values for each suggested boundary.
- Second, redundant segments are identified, and the most important segments are selected. Four partners provided ranked lists of redundant and selected segments, and those lists were fused to produce the final ranked list of common selected segments.
- Third, the final video summary is constructed by accelerating and concatenating the top-ranked selected video segments, according to the maximum allowed duration.

2   Results

Two runs were submitted to the TrecVid evaluation, using two different playback acceleration strategies [2]. The results, as determined by the formal TrecVid evaluation, are shown in the following table:

	K-Space 1	K-Space 2	Baseline	Mean
Fraction of inclusions	0.29	0.27	0.83	0.46
Junk (5 is best)	3.32	3.38	2.66	3.23
Duplicates (5 is best)	3.70	3.60	2.02	3.35
Pleasant (5 is best)	2.90	2.68	1.44	2.79

When compared with the TrecVid baseline system and mean of participating groups, these results show a low performance on the fraction of inclusions, probably due to the limited amount of training data that we could use to tune the fusion methods. The detection of junk and duplicates seem to be quite good, and the final acceleration is also reasonable for users to view.

The current approach is a first attempt at combining several analysis techniques for summarization. We hope that by improving the training phase of the system, we will be able to better benefit of the advantages of the various technical approaches provided by the different partners and that the combination will produce more accurate results.

References

1. K-Space Network of Excellence, http://www.k-space.eu/
2. Dumont, E., Merialdo, B., Essid, S., Bailer, W., Rehatschek, H., Byrne, D., Bredin, H., O'Connor, N., Jones, G., Smeaton, A., Haller, M., Kruitz, A., Sikora, T., Piatrik, T., Rushes Video Summarization Using a Collaborative Approach, ACM Multimedia TrecVid Summarization Workshop, Vancouver, Canada, October 27-31, 2008