Text Mining Support in Semantic Annotation and Indexing of Multimedia Data

Jan Nemrava[1], David Sadlier[2], Paul Buitelaar[3], Thierry Declerck[3]

[1] Engineering Group at the University of Economics in Prague (VSE)
W. Churchill Sq 4, 130 67 Prague, Czech Republic
nemrava@vse.cz
[2] Centre for Digital Video Processing at Dublin City University (DCU)
sadlierd@eeng.dcu.ie
[3] German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
{declerck, Paul.Buitelaar}@dfki.de

Abstract

This short paper is describing a demonstrator that is complementing the paper "Towards Cross-Media Feature Extraction" in these proceedings. The demo is exemplifying the use of textual resources, out of which semantic information can be extracted, for supporting the semantic annotation and indexing of associated video material in the soccer domain. Entities and events extracted from textual data are marked-up with semantic classes derived from an ontology modeling the soccer domain. We show further how extracted Audio-Video features by video analysis can be taken into account for additional annotation of specific soccer event types, and how those different types of annotation can be combined.

Introduction

Audio-visual (A/V) material does not support per se the use of text-based search within multimedia content in order, for example, to retrieve a specific situation of interest (highlight). But many A/V consumers would like to formulate their wishes/needs on seeing a specific A/V sequence using natural language queries. This generates a task for Multimedia Indexing consisting in adding some textual content to A/V content and semantically annotating (beyond keywords) the A/V stream with additional information, which is extracted from resources that are related to the A/V stream. This information can be added manually to the video content, or automatically by using natural language analysis of the related textual sources.

Optimally the information extracted from complementary sources can be combined with the features extracted directly from the A/V stream. This integration work is a research being addressed in the Network of Excellence K-Space[1], which is dedicated to the semantic inference for semi-automatic annotation and retrieval of multimedia content. The aim is to narrow the gap between A/V content descriptors that can be computed automatically and the richness and subjectivity of semantics in high-level human interpretations of entities and events present in the audiovisual content. This is done by exploiting and linking important multidisciplinary aspects of multimedia knowledge analysis, including *Content-based multimedia analysis*, *Knowledge extraction* and *Semantic multimedia*.

A good example of this research programme in K-Space is given with the sports domain. Most research in sports video analysis focuses on event recognition and classification based on the extraction of low-level features and is limited to a very small number of different event types, e.g. 'scoring-event'. But there are vast textual data that can serve as a valuable source for more fine-grained event recognition and classification, along the line of an ontology modelling the soccer domain. With a Multimedia ontology infrastructure in the background, in which an ontology for linguistic descriptors has been integrated, the integration of features extracted from text and from audio-video analysis seems to be possible, at least within certain domains, which can offer a good balance of available data in the form of A/V content and textual sources. We expect thus the presented work to be not confined to soccer only.

The kind of complementary resources we have been looking for the demonstrator presented here is given by textual documents that are external to the A/V stream. While it would be very interesting to include textual codes present in the video stream or speech data available in the audio-stream, technologies for extracting meaningful

[1] The research described in this paper is supported by the European Commission, contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space., see www.k-space.eu.

interpretation out of this textual and speech data is still not accurate enough. But, as in the case of soccer good quality complementary textual data, closely related to the A/V stream, is available in various formats, we could start our work on the possible integration of textual and audio-video features on a good empirical base.

Cross-Media Feature Extraction in the Soccer Domain

In the demo presented here, we are dealing for the textual analysis with structured web data (tables with teams, player names, goals, substitutions, etc.) and unstructured textual web data (minute-by-minute match reports). Information on entities and events extracted from those documents are stored as instances of semantic classes derived from SWIntO, the SmartWeb soccer ontology [2], by use of the soccer instantiation of the IE system SProUT [3], which was also developed in the SmartWeb project [4]. The temporal information associated with such instances extracted from the textual sources is normally expressed in minutes (the point in time in which events are reported). Those instances can then be used for the semantic indexing and retrieval of soccer videos.

This approach was followed in some respect earlier in the MUMIS project [1], and our work in K-Space builds on results of this project, but additionally aims at the extraction of A/V features. Feature detectors applied in the demonstrator are *motion activity measure, crowd detection, speech-band audio activity, presence/absence tracking, field-line*, and *close-up*, and are combined by means SVM (Support Vector Machine). We consider those features as being relevant for specific soccer event types, e.g. a CornerKick event will have a specific value for the field-line feature (EndLine), a ScoreGoal event will have a high value for the audio-pitch feature, etc.

Through the temporal alignment of the primary A/V data (soccer videos), which are available in "seconds" with the complementary resources, which are accounted for in "minutes" or even larger temporal interval, these extracted and semantically organized events can act as indicators for video segment extraction and semantic classification.

Conclusions

We described a demonstrator (depicted in figure 1 below) that implements an approach using the results of ontology driven natural language analysis of textual resources that are complementary to soccer A/V streams for supporting their semantic indexing and retrieval. We further presented event detection in the video based on general A/V detectors. Features extracted from both analysis types are temporally aligned, and so the video indexing and retrieval can be refined with semantics extracted for the complementary resources. As such extraction of A/V features is supported by textual evidence, and as textual extraction can possibly also be supported by evidence of features extracted from the A/V stream, our demonstrator is showing a good example of an implementation of research in the field of "cross-media feature extraction".

Short References

[1] http://www.dfki.de/pas/f2w.cgi?ltc/mumis-e

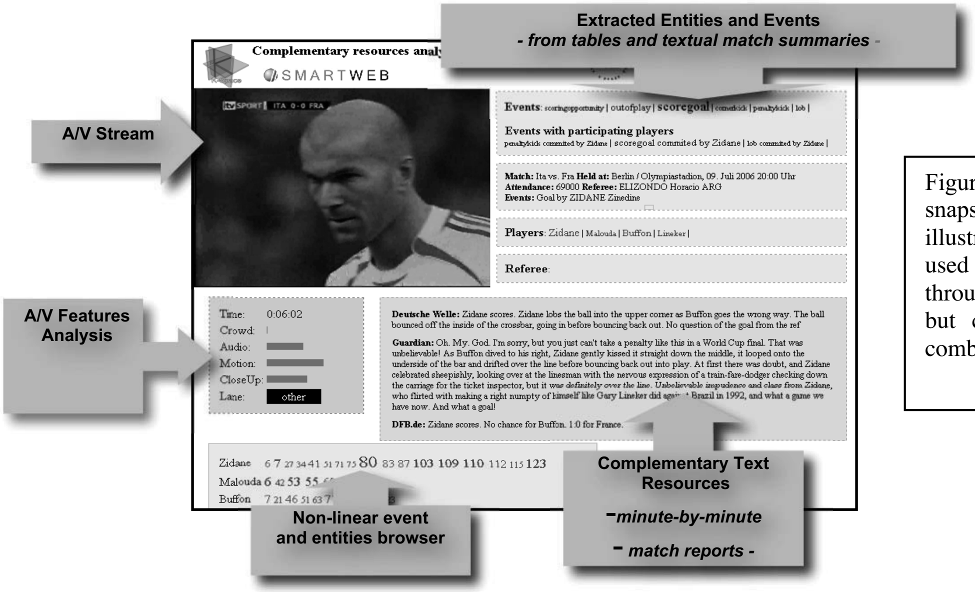
[2] http://www.dfki.de/sw-lt/olp2_dataset/

[3] http://sprout.dfki.de/

[4] http://smartweb.dfki.de/

[5] Sadlier D and O'Connor N.: Event Detection in Field Sports Video using Audio-Visual Features and a Support Vector Machine. IEEE Transactions on Circuits and Systems for Video Technology, Oct 2005

[6] D. Yow, B-L. Yeo, M. Yeung and B. Liu, "Analysis and presentation of soccer highlights from digital video," Proc. *Asian Conference on Computer Vision1995*, Singapore.



Extracted Entities and Events
- from tables and textual match summaries -
A/V Stream
A/V Features Analysis
Complementary Text Resources
-minute-by-minute
- match reports -
Non-linear event and entities browser
Complementary resources analy
SMARTWEB
SPORT ITA 0-0 FRA

Events: scoringopportunity | outofplay | **scoregoal** | cornerkick | penaltykick | lob |

Events with participating players
penaltykick commited by Zidane | scoregoal commited by Zidane | lob commited by Zidane |

Match: Ita vs. Fra **Held at:** Berlin / Olympiastadion, 09. Juli 2006 20:00 Uhr
Attendance: 69000 **Referee:** ELIZONDO Horacio ARG
Events: Goal by ZIDANE Zinedine

Players: Zidane | Malouda | Buffon | Lineker |

Referee:

Time: 0:06:02
Crowd:
Audio:
Motion:
CloseUp:
Lane: other

Deutsche Welle: Zidane scores. Zidane lobs the ball into the upper corner as Buffon goes the wrong way. The ball bounced off the inside of the crossbar, going in before bouncing back out. No question of the goal from the ref

Guardian: Oh. My. God. I'm sorry, but you just can't take a penalty like this in a World Cup final. That was unbelievable! As Buffon dived to his right, Zidane gently kissed it straight down the middle, it looped onto the underside of the bar and drifted over the line before bouncing back out into play. At first there was doubt, and Zidane celebrated sheepishly, looking over at the linesman with the nervous expression of a train-fare-dodger checking down the carriage for the ticket inspector, but it *was definitely over the line. Unbelievable impudence and class from Zidane*, who flirted with making a right numpty of himself like Gary Lineker did against Brazil in 1992, and what a game we have now. And what a goal!

DFB.de: Zidane scores. No chance for Buffon. 1:0 for France.

Zidane 6 7 27 34 41 51 71 75 80 83 87 103 109 110 112 115 123
Malouda 6 42 53 55
Buffon 7 21 46 51 63 7

Figure 1: The figure on the left shows a snapshot of our SMIL-based demo and illustrates the approach and the resources used in our work The user can navigate through a cross media information display, but can also query using different and combined media indicators.