Video Semantics and the Sensor Web

Alan F. Smeaton

CLARITY: Centre for Sensor Web Technologies
Dublin City University
Glasnevin, Dublin 9, Ireland
{alan.smeaton}@dcu.ie

The most widespread way in which content-based access to video information is supported is through using a combination of video metadata (date, time, format, etc.) and user-generated description (user tags, ratings, reviews, etc.). This has had widespread usage and is the basis for navigation through video archives in systems such as YouTube, Open Video and the Internet Archive. However, there are limitations with this such as vocabulary issues, and authentication across the users who annotate content. Our work has concentrated on addressing *content-based* video retrieval, the kind of direct retrieval we routinely perform on text, retrieving web pages or blog posts based primarily on actual content. One premise for our work is that we can extract descriptions of content reliably. We can easily extract low-level features such as colour, texture and shapes from the visual aspects of video (the picture) and from these we can compute visual similarity. In a video retrieval system this manifests as being able to find video shots based on the visual similarity between one or more query images and each of the keyframes in the video. Once again, however, this also has limits, such as having to locate query images to start the query, and the fact that a keyframe is just one moment of a video shot which may not be representative.

More relevant to this talk is the challenge of automatically identifying semantic features from video and the talk will present a snapshot of the current approaches taken to identifying the presence or absence of groups of semantic features, in video. The annual TRECVid activity has been benchmarking the effectiveness of various approaches since 2001 and we will examine what is the performance of these detectors, what are the trends in this area, and what is the state of the art in this. We will discover that the performance of individual detectors varies widely depending on the nature of the semantic feature, the quality of training data and its dependence on other detectors. Current work addresses the inter-dependence among semantic features, supported by the use of ontologies, and also how the relatively poor quality of detection of these semantic features when taken in isolation, can be ameliorated by orchestrating them to work together. There is a strong analogy between this and the way that sensors (environmental, physiological, etc.) which make up the sensor web, can also have poor accuracy levels when used in isolation but whose individual performances can be improved when used in combination.

Acknowledgment: The author acknowledges the support of Science Foundation Ireland under grant number 07/CE/I1147.