ANALYZING VIDEO CONCEPT DETECTORS VISUALLY

Cees G.M. Snoek, Richard van Balen, Dennis C. Koelma, Arnold W.M. Smeulders, Marcel Worring

ISLA, University of Amsterdam
Kruislaan 403
1098 SJ Amsterdam, The Netherlands
http://www.mediamill.nl

ABSTRACT

In this demonstration we showcase an interactive analysis tool for researchers working on concept-based video retrieval. By visualizing intermediate concept detection analysis stages, the tool aids in understanding the success and failure of video concept detection methods. We demonstrate the tool on the domain of pop concert video.

Index Terms— video concept detection, analysis tool;

1. INTRODUCTION

The key-objective in video retrieval is to provide access at the semantic level, by describing the combination of concepts appearing in the visual content. Research aiming to fulfill this challenging goal suffers from the well known semantic gap. In an effort to bridge this semantic gap, the video retrieval community has been focusing its attention on an analysis method based on so called concept detectors. Concept detectors aim to provide semantic access by labeling visual content with descriptive terms like *parade*, *mobile phone*, and *desert*.

A basic video concept detector contains two stages. In the first stage the detector extracts features from the video data. Subsequently the extracted features together with labeled concept examples, form the input for a supervised learning stage. At training time this results in a model, when applied at testing time it yields a concept detection likelihood in the form of a probability. The probability serves as a ranking criterion, which allows us to evaluate concept detection performance

From a researchers perspective, the performance of a concept detector is dependent on many variables. With so many choices on offer it is hard to pinpoint why a concept detector succeeds or fails in a given video. Is it caused by the features used, the provided concept examples, the classifier? And, as video is an inherently temporal medium, does the observed effect appear only occasionally or is it a structural phenomenon? Getting a grip on all these variables is problematic.

In this demonstration we present an interactive tool that aids researchers in understanding the success and failure of video concept detectors. This is achieved by visualizing relevant data from all intermediate concept detection analysis stages.

2. THE DEMONSTRATION

To demonstrate the utility of our tool we use pop concert video from Fabchannel, who narrowcast live concerts from club venues in Amsterdam over the Internet. On a selection of Fabchannel video we extract standard MediaMill visual features [1] on the frame-level. For supervised learning we adopt the concert concepts defined in [2], and we employ an SVM with episode-constrained cross-validation [3] to obtain concept likelihoods. Naturally our analysis tool offers basic functionality like selecting and viewing the concert videos. In the demonstration we show that by offering an integrated perspective on several intermediate concept detector analysis stages, our interactive visual analysis tool supports researchers in understanding video concept detector behavior. We summarize the various visualization components of our interactive concept detection analysis tool in Table 1.

3. REFERENCES

[1] C.G.M. Snoek, et al., "The MediaMill TRECVID 2007 Semantic Video Search Engine," In *Proc. TRECVID Workshop*, Gaithersburg, MD, 2007.

[2] C.G.M. Snoek, et al., "The Role of Visual Content and Style for Concert Video Indexing," In *Proc. IEEE ICME*, pp. 252-255, Beijing, China, 2007.

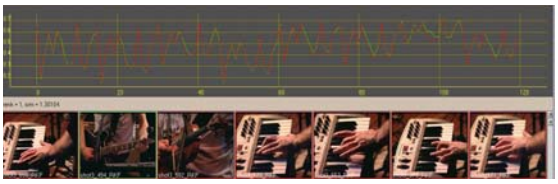
[3] J.C. van Gemert, et al., "The Influence of Cross-Validation on Video Classification Performance," In *Proc. ACM Multimedia*, pp. 695-698, Santa-Barbara, CA, 2006.

This research is supported by the BSIK MultimediaN project and the EU IST-CHORUS project. We are grateful to Bauke Freiburg from Fabchannel for providing us with the concert videos.

CONCEPT DETECTION ANALYSIS STAGE	VISUALIZATION COMPONENT

Feature Extraction

- Visual comparison of feature vectors aids in understanding why similar concepts are detected in highly dissimilar images, and why dissimilar concepts are detected in similar images,

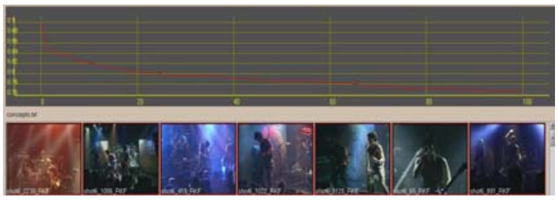


Manual Concept-based Labeling

- Visual comparison of annotated images aids in understanding the visual complexity of the concept under study,

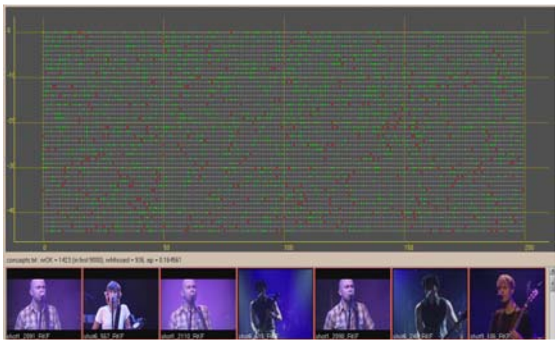
Supervised Machine Learning

- Visual comparison of concept detector likelihood scores aids in understanding concept detector reliability on the domain under study,



Performance Evaluation

- Visual comparison of detected concepts with a manual labeled ground truth aids in assessing overall concept detector effectiveness, and provides specific feedback on cases of success and failure,



Temporal Concept Behavior

- Visual comparison of (multiple) concept detector likelihood scores over multiple levels of video granularity aids in understanding temporal concept behavior and it offers insight in the question how uncertain concept combinations may describe video content;

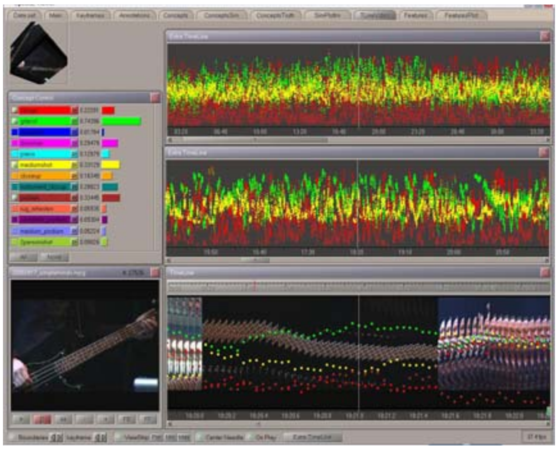


Table 1: *We demonstrate an interactive analysis tool that aids researchers in understanding concept detection performance by visualizing intermediate concept detector analysis stages, including (from top to bottom) feature extraction, manual labeling, supervised learning, evaluation, and temporal behavior.*