High-level Event Detection System Based on Discriminant Visual Concepts

Ioannis Tsampoulatidis, Nikolaos Gkalelis*, Anastasios Dimou, Vasileios Mezaris, Ioannis Kompatsiaris
Informatics and Telematics Institute
Centre for Research and Technology Hellas
6th Km Charilaou-Thermi Road, Thermi 57001, Greece
{itsam, gkalelis, dimou, bmezaris, ikom}@iti.gr

ABSTRACT

This paper demonstrates a new approach to detecting high-level events that may be depicted in images or video frames. Given a non-annotated content item, a large number of previously trained visual concept detectors are applied to it and their responses are used for representing the content item with a model vector in a high-dimensional concept space. Subsequently, an improved subclass discriminant analysis method is used for identifying a concept subspace within the aforementioned concept space, that is most appropriate for detecting and recognizing the target high-level events. In this subspace, the nearest neighbor rule is used for comparing the non-annotated content item with a few known example instances of the target events. The high-level events used as target events in the present version of the system are those defined for the TRECVID 2010 Multimedia Event Detection (MED) task.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Experimentation, Algorithms, Human Factors

1. INTRODUCTION

The automatic understanding of the semantic content of multimedia has long been identified as a major research goal towards the efficient manipulation of the content. Tasks such as image/video retrieval, video summarization, and surveillance applications are among those that can benefit from advances in this line of research, such as advances in the automatic detection of concepts (e.g. objects such as "sea", "person" etc.) that are depicted in the visual medium. Recently, the detection of the high-level events that are related with different multimedia content items, rather than just the detection of simpler concepts depicted in them, has started to receive significant attention by the research community [4]. The reasons behind this shift from purely concept-based to event-based understanding efforts lie in the significance of real-life events in the human-centered organization of media.

*N. Gkalelis is also with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK.

ICMR '11, April 17-20, Trento, Italy
ACM 978-1-4503-0336-1/11/04.

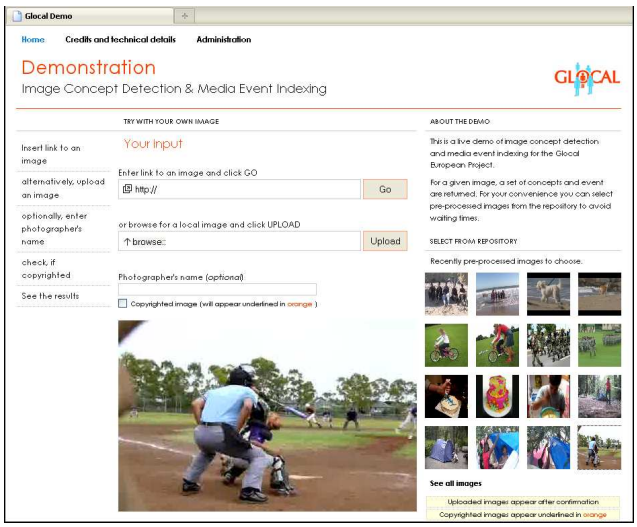

Figure 1: User interface of the event detection system.

2. EVENT DETECTION SYSTEM

In this work we demonstrate a system for detecting and recognizing high-level events in images or video keyframes. The simple web-based user interface that enables the interaction with the proposed system is shown in Fig.1. The user, after accessing the interface via the web, is allowed to submit images and view the results of event detection for each submitted image. For image submission, the user is presented with the option to either upload an image that resides locally in the user's computer, or specify a web address where the desired image can be found. Any file format (e.g. jpg, png, etc.) is acceptable in both cases. The results of event detection, obtained according to the method outlined in the following section, are then presented to the user as shown in Fig. 2.

Figure 2: Presentation of intermediate and final results of the event detection system (i.e., concept detection scores, using a horizontal bar for each of the 231 concepts; and the name of the one event among the target ones that was eventually detected, if any, in a yellow-highlighted field).

3. EVENT DETECTION METHOD

The event detection method that is used as part of the proposed system is a visual-model-vector based approach, where visual concept detectors are used to automatically describe a video sequence in a concept space. Visual model vectors were originally proposed for the task of image/video retrieval [6]. In the present system, we use a total of 231 visual concept detectors, which we trained on pre-existing concept-annotated datasets (i.e., 101 detectors trained on the MediaMill dataset [7], and 130 additional detectors trained on part of the TRECVID 2010 Semantic Indexing (SIN) Task dataset [5]), for forming the model vectors.

Following the representation of each image with a 231 element model vector, a novel Discriminant Analysis (DA) technique, Mixture Subclass Discriminant Analysis (MSDA) [2], is invoked. This is used for identifying the semantic concepts that best describe the target events, thus defining a discriminant concept subspace for each event. The MSDA method extends the recently proposed Subclass Discriminant Analysis (SDA) [8], to further improve the recognition accuracy and the degree of dimensionality reduction. The latter is an important factor for ensuring the timely completion of event detection. In the resulting discriminant subspace, the nearest neighbor classifier (NN) is used for recognizing one of the target events, if any, with the help of a few known example instances of the target events.

For the purpose of experimentation and demonstration, the target events and training data provided as part of the MED task of TRECVID 2010 are used. The target events include three high-level events, namely, "batting a run in" (which refers to a specific event during baseball games), "making a cake" and "assembling a shelter". The training data include a few tens of images depicting each of the three target events, and several more depicting other, unidentified, non-interesting events.

More algorithmic details on preliminary versions of the event detection method used in this work and also on specific elements of it, such as the improved subclass discriminant analysis technique, can be found in [1], [2], [3].

4. CONCLUSIONS

In this work, visual concept detectors were used for supporting the task of detecting high-level events in images, and a method based on using a relatively large number of such detectors together with an improved subclass discriminant analysis technique was proposed for leveraging the inaccuracy of each individual concept detector. Experimental results obtained in the context of TRECVID 2010 as well as on other existing datasets document the merits of the proposed approach.

5. ACKNOWLEDGMENTS

This work was supported by the European Commission under contract FP7-248984 GLOCAL.

6. REFERENCES

[1] N. Gkalelis, V. Mezaris, and I. Kompatsiaris. Automatic event-based indexing of multimedia content using a joint content-event model. In *Proc. ACM Multimedia 2010, Events in MultiMedia Workshop (EiMM10)*, Firenze, Italy, October 2010.

[2] N. Gkalelis, V. Mezaris, and I. Kompatsiaris. Mixture subclass discriminant analysis. *IEEE Signal Processing Letters*, 2011.

[3] A. Moumtzidou, A. Dimou, N. Gkalelis, S. Vrochidis, V. Mezaris, and I. Kompatsiaris. ITI-CERTH participation to TRECVID 2010. In *Proc. TRECVID 2010*, Gaithersburg, MD, USA, November 2010.

[4] A. Scherp, R. Jain, M. Kankanhalli, and V. Mezaris. Modeling, Detecting, and Processing Events in Multimedia. In *Proc. ACM Multimedia 2010*, pages 1739–1740, Firenze, Italy, October 2010.

[5] A. F. Smeaton, P. Over, and W. Kraaij. High-Level Feature Detection from Video in TRECVid: a 5-Year Retrospective of Achievements. In A. Divakaran, editor, *Multimedia Content Analysis, Theory and Applications*, pages 151–174. Springer Verlag, Berlin, 2009.

[6] J. Smith, M. Naphade, and A. Natsev. Multimedia semantic indexing using model vectors. In *Proc. IEEE ICME*, pages 445–448, Baltimore, MD, USA, July 2003.

[7] C. Snoek, M. Worring, J. van Gemert, J.-M. Geusebroek, and A. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proc. ACM Multimedia*, pages 421–430, Santa Barbara, USA, October 2006.

[8] M. Zhu and A. Martinez. Subclass discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(8):1274–1286, August 2006.