Visor: Video Surveillance Online Repository

Roberto Vezzani and Rita Cucchiara

Imagelab – Dipartimento di Ingegneria dell'Informazione
University of Modena and Reggio Emilia, Italy

1. Introduction

Aim of the Visor Project [1] is to gather and make freely available a repository of surveillance and video footages for the research community on pattern recognition and multimedia retrieval. The goal is to create an open forum and a free repository to exchange, compare and discuss results of many problems in video surveillance and retrieval.
Together with the videos, the repository contains metadata annotation, both manually annotated as ground-truth and automatically obtained by video surveillance systems. Annotation refers to a large ontology of concepts on surveillance and security related objects and events. The ontology has been defined including concepts from LSCOM and MediaMill ontologies. As well as videos and annotations, Visor provides tools for enriching the ontology, annotating new videos, searching by textual queries, composing and downloading videos.

2. Video Surveillance Concept List

To ensure interoperability between users a standard annotation format has been defined together with the structure of the knowledge base. The knowledge which could be extracted from video surveillance clips is structured as a simple "concept list": this taxonomy is a basic form of ontology where concepts are hierarchically structured and univocally defined. The concept list can by dynamically enriched by users under the supervision of the Visor manager to ensure the homogeneity and the uniqueness. The goal is to create a very large concept list avoiding synonymy and polysemy drawbacks.

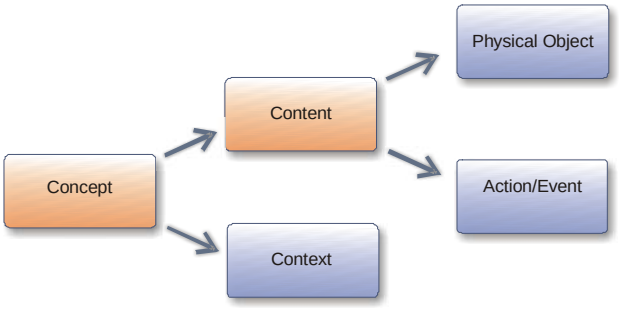

Fig. 1: Hierarchical taxonomy of the video concept categories

We defined a basic taxonomy to classify the video shapes, objects and highlights meaningful in a surveillance environment (see Fig. 1). A "concept" can describe either the *context* of the video (e.g., *indoor, traffic surveillance, sunny day*), or the content which can be a *physical object* characterizing or present in the scene (e.g., *building, person, animal*) or a detectable *action/event* (e.g., *falls, explosion, interaction between people*).
The defined concepts can be differently related with the time space. Thus, we can introduce a time based taxonomy of the concepts. A concept can be associated to the whole video (e.g.: *indoor, outdoor*), to a clip/temporal interval (e.g., *person in the scene*), or to a single frame/instant (e.g., *explosion, person entering the scene*).
A first reference concept list has been obtained as a subset of two different predefined sets, respectively the 101-concept list of UvA[2] and LSCOM[3]. Since these lists have been defined for generic contexts, only a subset of the reported concepts have been elicited for video surveillance. Moreover, UvA and LSCOM lists are key-frame based only and are not enough to describe activities and events. An extension of the base LSCOM list have been considered (LSCOM Revised Event/Activity Annotations: video-based re-labeling of 24 LSCOM concepts [4]), but it is still limited. Thus, we have collected and reported other concepts we are interesting on; most of them are defined at a very high abstraction level. Actually, a preliminary list of more than 100 surveillance concepts has been defined.
With reference to the taxonomy of Fig. 1, the video surveillance concepts can belong to three semantically different categories. The Visor ontology is structured in several classes, each of them belonging to one category as reported in Table 1. A video annotation can be considered as a set of instances of these classes; for each instance a list of related concepts are assigned. Some of them directly describe the nature of the instance, or, in other words, they are connected to the entity with a "IS-A" relation (e.g., concepts like *man*, *woman*, *baby*, *terrorist* can be assigned to instances of the *Person* class). Other concepts, instead, describe some characteristics of the instance, in a "HAS-A" relation with it (e.g., the *contour*, the *color*, the *position* can be descriptive features of *FixedObject* instances).

Class	Category
Person	*PhysicalObject*
BodyPart	*PhysicalObject*
GroupOfPeople	*PhysicalObject*
FixedObject	*PhysicalObject*
MobileObject	*PhysicalObject*
ActionByAPerson	*Action/Event*
ActionByPeople	*Action/Event*
ObjectEvent	*Action/Event*
GenericEvent	*Action/Event*
Video	*Context*
Clip	*Context*
Location	*Context*

Table 1: Set of surveillance classes

3. Annotation format

The native annotation format supported by VISOR is Viper[5], developed at the University of Maryland. The selection of this annotation format is due to several requirements that Viper satisfies: it is flexible, the list of concepts is customizable; it is widespread avoiding the difficulties to share a new custom format; it is clear and easy to use, self containing since the description of the annotation data is included together with the data. Finally, it is possible to achieve a frame level annotation that is more appropriate than the clip level annotation adopted by other tools.

4. Web Interface

The Visor web interface has been developed in order to share the videos and the annotation contents. Some screenshots of the web interface are shown in Fig. 2. Visor supports multiple video formats, search by keywords, by video metadata (e.g., author, creation date, ...), by camera information and parameters (e.g., camera type, motion, IR, omni-directional, calibration).
Three modalities have been implemented to allow video access in different ways: video preview, based on a compressed stream, single screenshot (a representative frame of the entire video) or a summary view, in which clip level screenshots are reported. For each video, a set of annotations are provided, both ground truth and automatic annotations. The web interface allows to download the entire annotation as well as a subset of the annotation fields, filtering by frame number, descriptor or single attribute. The annotation can be exported in the VIPER format; an MPEG7 format exportation module is under development.

5. Forum

Another important aspect for a research community is the information exchange and the opportunity to share opinions, requests, comments about the videos and the annotations, and so on. Thus, the online portal of Visor includes a forum in which one topic for each video, generic topics on video surveillance, and topics on VISOR (e.g., call for videos) are already active. In addition, each registered user can create his own topics.

6. Conclusion and future work

Visor is a repository of annotated video sequences related to surveillance applications. A suitable ontology for surveillance domains has been defined in order to assure a better and easier interoperability among users. The annotation are in the Viper XML format but an MPEG7 exportation module is under development.

7. Acknowledgements

This work was supported by the project VidiVideo (Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts), funded by EC VI Framework programme.

References

[1] http://imagelab.ing.unimore.it/visor
[2] Cees G. M. Snoek, Marcel Worring, Jan C van Gemert, Jan Mark Geusebroek, and Arnold W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In Proc ACM-Multimedia. ACM-Press, 2006
[3] M. R. Naphade, L. Kennedy, J. R. Kender, S.-F. Chang, J. R. Smith, P. Over, and A. Hauptmann, “A Light Scale Concept Ontology for Multimedia Understanding for TRECVID 2005,” IBM Research Technical Report, 2005
[4] Lyndon Kennedy, Revision of LSCOM Event/Activity Annotations, DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia, Columbia University ADVENT Technical Report #221-2006-7 , December 2006.
[5] http://viper-toolkit.sourceforge.net/

Fig. 2: Screenshots of the VISOR web interface