Smoke Detection in Video Surveillance: the Use of ViSOR (Video Surveillance On-line Repository)

Roberto Vezzani, Simone Calderara, Paolo Piccinini, Rita Cucchiara
Department of Information Engineering
University of Modena and Reggio Emilia
Via Vignolese 905/b, 41100 Modena, Italy
{roberto.vezzani,simone.calderara,paolo.piccinini,rita.cucchiara}@unimore.it

ABSTRACT

ViSOR (Video Surveillance Online Repository) is a large video repository, designed for containing annotated video surveillance footages, comparing annotations, evaluating system performance, and performing retrieval tasks. The web interface allows video browse, query by annotated concepts or by keywords, compressed video preview, media download and upload. The repository contains metadata annotations, both manually created ground-truth data and automatically obtained outputs of particular systems. An example of application is the collection of videos and annotations for smoke detection, an important video surveillance task. In this paper we present the architecture of ViSOR, the build-in surveillance ontology which integrates many concepts, also coming from LSCOM, and MediaMill, the annotation tools and the visualization of results for performance evaluation. The annotation is obtained with an automatic smoke detection system, capable to detect people, moving objects, and smoke in real-time.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries—*Collection, Dissemination, system issues*

General Terms

Design, Management, Performance

Keywords

Video repository, video surveillance ontology, annotation, smoke detection

1. INTRODUCTION

In the research activity for knowledge extraction, automatic annotation and content-based video retrieval a critical role is always the performance evaluation: especially in domain specific video repositories, the research communities needs ground truth data, available databases with annotations, tools for annotate media and make comparisons. In video surveillance some open source tools, such as Viper-GT and Viper-PE [15] are growing their popularity since they constitute an interoperable platform to manually select concepts and events in video, generate ground truth and annotate videos into XML files. The Viper annotation format [3] is widely exploited by available databases of videos which are created in contexts of workshops and conferences like the PETS workshop series[17] or the VSSN workshops of the ACM Multimedia Conference [16] and in the ones that become available from some European or national projects such as I-Lids[1] and Etiseo [9]. Some examples of available datasets are reported in the table of Figure 1. Most of these datasets have two main drawbacks. The first is their narrow focus on few specific problems of computer vision and pattern recognition. The PETS datasets, for instance, have been deeply exploited in some applications but they have been proposed within their a-priori annotation with the aim of coping a single or few video surveillance problems. The second limitation is the lack of user interaction; for example, user cannot share their own annotation data, or grow the dataset with other videos, or comment them, and so on. Moreover, the defined ontology is normally not available, and there are not graphical tools or querying systems to select only the subset of videos useful for a given application.

The Video Surveillance Online Repository (ViSOR) for annotation retrieval has been conceived to meet these needs. It has been designed and is under development in the context of the Vidi-Video European projects of the VI Frame Program. First aim of ViSOR is to gather and make freely available surveillance video footages for the research community on pattern recognition and multimedia retrieval. At the same time, our goal is to create an open forum and a interactive repository to exchange, compare and discuss results of many problems in video surveillance and retrieval.

Together with the videos, ViSOR defines an ontology for metadata annotations, both manually provided 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, defined including concepts from LSCOM and MediaMill ontologies. Moreover, ViSOR provides tools for enriching the ontology, annotating new videos, searching by textual queries, composing and downloading videos.

In particular, in this paper we present a section of ViSOR specialized to smoke detection algorithms. Smoke detection in video-surveillance systems is still an open challenge for

CIVR'08, July 7–9, 2008, Niagara Falls, Ontario, Canada.

computer vision and pattern recognition communities. It concerns the definition of robust approaches to detect, as soon as possible, spring and fast propagation of smoke possibly due to explosions, fires or special environmental conditions. Smoke detection module can enrich standard video surveillance systems for both indoor and outdoor monitoring. Moreover, detecting smoke by visual cues could allow fast and reactive alarms also in some specific situations, where smoke is growing in unconventional directions, so that the time-to-alarm of normal sensors could become unacceptable. In addition it is a typical example of unusual event that could be the target of a video retrieval system. In [10], for instance an example of inter-operable software tool for the analysis of event detection system was presented.

The video analysis tasks for smoke detection are not trivial due to the variability of shape, motion and texture patterns of smoke, which appearance is dependent on the luminance conditions, the background manifolds and colors of the scene. Since smoke modifies the visual cues of the scene, typically background suppression techniques are adopted, followed by validation/classification tasks. The smoke identification becomes more challenging in presence of other moving objects and shadows and whenever the background is variable as well.

In this paper we describe the ViSOR project, the video surveillance ontology and the web interface for analyzing and querying the results of automatic event detection systems. ViSOR architecture is very general, can be enriched with general-purpose annotation and visualization tools. Here, we present a graphical tool for showing results of people and smoke detection is presented. ViSOR is an open platform that will be useful for the video surveillance and multimedia community in order to collect free data, and comparing tools and systems of annotation and retrieval.

2. VIDEO SURVEILLANCE ONTOLOGY IN VISOR

Some proposals of ontologies for event detection in surveillance have been carried out. An example is the ontology defined in the Etiseo project [9] or the result of the "Challenge Project on Video Event Taxonomy" sponsored by the Advanced Research and Development Activity (ARDA) [8]. In [4] a Video Event Representation Language (VERL) is presented which describes an event ontology, associated with Video Event Markup Language (VEML) for event instance annotation. Viper-gt [15], instead, is a very spread graphical tool for manual annotation of objects and frame-based events, exploited in video-surveillance community.

Here we start from the Viper framework and propose an open simple ontology 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 moderator to ensure the homogeneity and the uniqueness. The goal is to create a very large concept list avoiding synonymy and polysemy drawbacks.

2.1 Video Surveillance Concepts

We defined a basic taxonomy to classify video shapes, objects and highlights meaningful in a surveillance environment. 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* occurring (e.g., falls, explosion, interaction between people).

The defined concepts can be differently related with the time space. Thus, we defined a time based taxonomy of the concepts depending on its span, e.g. the time interval during which the object is visible or the event/action is occurring. 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[12] and LSCOM[7]. 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 [5]), but only few concepts refer to surveillance. Thus, we have collected and reported other interesting concepts; most of them are defined at a very high abstraction level. Actually, a preliminary list of more than 100 surveillance concepts has been defined.

The video surveillance concepts can belong to three semantically different categories (*Physical Object, Action/Event, Context*). More precisely, the ViSOR ontology is structured in several classes, each of them belonging to one of the previously defined categories as reported in Table 2. 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, i.e., they are connected to the entity with a "IS-A" relation (e.g., concepts like man, woman, baby, terrorist can be a sort of specialization of the "person" class and thus they can be use to describe instances of that class). Other concepts, instead, describe some characteristics or properties of the instance, in a "HAS-A" relation with it (e.g., the contour, the color, the position, the bounding box can be descriptive features of *FixedObject* instances).

Specialization relations are always *static*, i.e, they do not change during time; for example, a person can be a man or a woman, but reasonably it cannot switch between them during the video clip. Differently, some "HAS-A" relation can be *dynamic*; for example, the position and the color of the person can be different frame by frame. Thus, we have distinguished the "HAS-A" concepts in *static* and *dynamic*. In other words, the appellation *dynamic* indicates that the concept has a dynamic evolution of some of its visual properties, and thus may be recognized performing an analysis that goes beyond a single key-frame description, or may provide more information if this evolution is taken into account. A complete list of the video surveillance concepts can be directly downloaded from the ViSOR portal.

2.2 Smoke detection concepts

Traditionally, a smoke detector or smoke alarm is a device that detects smoke and issues an alarm to alert nearby people that there is a potential fire. A computer-vision smoke detector should perform the same task, i.e. identify the pres-

Dataset		Website	Topics	Ground-Truth	Size
CANDELA		http://www.multitel.be/~va/candela/	Indoor left-luggage and traffic monitoring on road intersection	no	16 indoor
Etiseo		http://www-sop.inria.fr/orion/ETISEO/	Object Detection, Object Localization, Object Tracking, Object Classification.	yes	86 video clips
i-Lids (AVSS 2007)		ftp://motinas.elec.qmul.ac.uk/pub/iLids/	Stopped vehicles and abandonded luggage	yes	14 sequences
ObjectVideo Virtual Video		http://development.objectvideo.com/	Tool to generate virtual video sequences for surveillance purposes.	yes	-
PETS	2001	http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html	Outdoor people and vehicle tracking	yes	5 sequences
	2002	http://www.cvg.cs.rdg.ac.uk/PETS2002/pets2002-db.html	Indoor people tracking (and counting)	yes	6 sequences
	2004	http://www-prima.inrialpes.fr/PETS04/caviar_data.html	People tracking and activity recognition	yes	28 sequences, 6 scenarios
	2006	http://pets2006.net/	Surveillance of public spaces, detection of left luggages	yes	7 datasets (4 camera views each one)
	2007	http://pets2007.net/	Multisensor sequences containing loitering, attended luggage removal (theft), and unattended luggage	yes	8 datasets (4 camera views each one)
SELCAT		http://www.multitel.be/~va/selcat/	Level crossing monitoring for stopped vehicles detection.	yes	8 sequences
SPEVI		http://www.spevi.org	Face detection and tracking	partial	10 sequences
Traffic datasets by Institut für Algorithmen und Kognitive Systeme		http://i21www.ira.uka.de/image_sequences/	Traffic surveillance in particular on road intersections	no	14 sequences
VISOR		http://imagelab.ing.unimore.it/visor	Indoor and outdoor surveillance sequences; annotation data for object detection, tracking, events, and much more.	yes	65 sequences at 12/12/2007 (in progress)
VSSN		http://imagelab.ing.unimore.it/vssn06/	background subtraction competition	no	7 sequences

Figure 1: Available surveillance datasets

Person			
"Is-a" Concepts			
Name	Definition	Type	Dynamic
Adult	Shots showing a person over the age of 18 (LSCOM #181)	True/False	-
Aggressor	(LSCOM #461)	True/False	-
Baby	images of babies (children that are too young to walk) (LSCOM #247)	True/False	-
Boy	One or more male children. (LSCOM #183)	True/False	-
Child	images of children (LSCOM #273)	True/False	-
Civilian_Person	One or more persons not in the armed services or police force. (LSCOM #105)	True/False	-
Female	(LSCOM #21)	True/False	-
Girl	One or more female children. (LSCOM #184)	True/False	-
Male	(LSCOM #17)	True/False	-
Person	Shots depicting a person. The face may be partially visible (LSCOM #217)	True/False	-
Police/security	(LSCOM #42)	True/False	-
"Has-a" Concepts			
Position_BBOX	Bounding box containing the person	rectangle	True
PositionBar	2D Position of the gravity center	point	True
Contour	Person's Contour	polygon	True
IDPerson	Application defined ID	integer	False
RealHeight	Real height of the person	float	False
PersonName	Person's Name	string	False
Mobile Object			
"Is-a" Concepts			
Bicycle	A person riding a bicycle. (LSCOM #197)	True/False	-
Bird	(LSCOM #79)	True/False	-
Bus	Shots of a bus (LSCOM #227)	True/False	-
Car	Shots of a car (LSCOM #221)	True/False	-
Chair	(LSCOM #56)	True/False	-
Smoke	Shots with smoke present. (LSCOM #161)	True/False	-
"Has-a" Concepts			
Position_BBOX	Bounding box containing the object	rectangle	True
PositionBar	2D Position of the gravity center	point	True
Contour	Contour of the object	polygon	True

Table 1: Excerpts taken from *Person* and *Mobile Object* concept lists.

ence of smoke generating an alarm. Additionally to the detection, a computer vision system can enrich the knowledge with visual characteristics of the smoke, for example with its position and its "size" (i.e., diffusion of the smoke in the scene). Moreover, a video surveillance system can provide information about people interacting with the smoke or people have been around the area where the smoke come from.

Taking into account the requirements of a smoke detection system above mentioned, we have defined a Smoke detection concept list containing:

- the *smoke* concept, which is considered as an "Is-A" attribute of the mobile object class;
- geometrical features of the smoke, like the *position*, the *contour*, and the *bounding box*;
- *person concepts* that can be used to describe people interacting with the smoke or being where it comes from (both 'Is-A" and "Has-A" attributes).

A list of some interesting Video Surveillance concepts defined in ViSOR and containing also concepts related to smoke detection is reported in Table 1; this list has been extracted from the video surveillance concept list available on the ViSOR web portal [14].

2.3 The annotation format

The native annotation format supported by ViSOR is Viper[3], developed at the University of Maryland. The choice of this annotation format has been made 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 (e.g., it is used by *Pets* and *Etiseo*); it is clear and easy to use, self containing since the description of the annotation data is included together with the data. Differently from other existing tools working only on textual annotation, a set of data types which can be used for annotate has been defined (see Table 2.3).

Moreover, an annotation tool has already been developed by the same authors of the standard (namely, ViPER-GT [15]). Finally, it is possible to achieve a frame level annotation that is more appropriate than the clip level annotation adopted by other tools.

The annotation data is stored as a set of records. Each record, called *descriptor*, annotates an associated range of frames with a set of attributes. To inform applications of the types of descriptors used to create the data file and the

Class	Category
1. Person	PhysicalObject
2. BodyPart	PhysicalObject
3. GroupOfPeople	PhysicalObject
4. FixedObject	PhysicalObject
5. MobileObject	PhysicalObject
6. ActionByAPerson	Action/Event
7. ActionByPeople	Action/Event
8. ObjectEvent	Action/Event
9. GenericEvent	Action/Event
10. Video	Context
11. Clip	Context
12. Location	Context

Table 2: Set of surveillance classes

Data Type	Description
bbox	A bounding box; it is a rectangle on the image.
bvalue	A Boolean value: either "true" or "false".
circle	A circle, in terms of center point and radius.
ellipse	An ellipse, in terms of its bounding box.
fvalue	A floating point number.
lvalue	An enumeration type. In the config part the list of allowed values must be defined.
obox	An oriented bounding box.
point	Some specific pixel in the image.
polygon	A polygon or polyline, given as a list of points.
relation	A set of object identification numbers to a certain type of descriptor.
svalue	A string value. Remember that strings must be xml-escaped.

Table 3: Viper Data types

data-types of the associated attributes, users must provide configuration information at the beginning of file. To this aim, Viper files are basically composed by two sections; the first one is called *config part* and explicitly outlines all possible descriptors in the viper file. It defines each descriptor type by name and lists all attributes for each descriptor. From the ViSOR portal a predefined *config file* containing the video surveillance concept list described in the previous section can be obtained. The second section of each Viper file, namely *data part*, contains instances of the descriptors defined in the *config part*. For each instance, the frame span (i.e., range) of the descriptor visibility together with a list of attributes values are reported. For more details refer to the Viper manual [15] or directly to the ViSOR portal [14].

3. FOREGROUND OBJECT CLASSIFICATION FOR SMOKE DETECTION

The main scope of this section is the overview of the whole system developed for smoke detection. More details can be found in [11]. The presented system is an ensemble of different modules as depicted in Fig.2. The proposed model

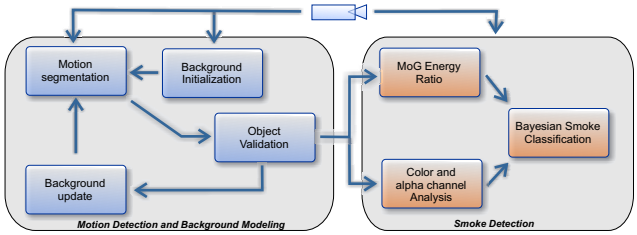

Figure 2: Block diagram of the smoke detection system.

evaluates the joint contribution coming from the gray level image energy and color intensity attenuation to classify an object as possible smoke. We assume that when smoke grows

and propagates in the scene its image energy is attenuated by the blurring effect of smoke diffusion.
We firstly detect possible candidate objects by means of a motion segmentation algorithm[2]. When a new foreground object is detected we analyze its energy using the Wavelet Transform coefficients and evaluate its temporal evolution. The color properties of the object are analyzed accordingly to a smoke reference color model to detect if color changes in the scene are due to a natural variation or not. The input image is then divided in blocks of fixed sized and each block is evaluated separately. Finally a Bayesian approach detect whether a foreground object is smoke.

3.1 Energy analysis using the direct wavelet transform

An efficient way to evaluate the energy variation of an intensity image is the discrete wavelet transform DWT [6].
The DWT is obtained convolving the image signal with several banks of filters obtaining a multi-resolution decomposition of the image. Given the input image I_t the decomposition produces four sub-images, namely the compressed version of the original image C_t, the horizontal coefficient image H_t, the vertical coefficient image V_t and the diagonal coefficient image D_t. An example decomposition is computed with the algorithm proposed in [6] is shown in Fig. 3

Figure 3: Example of discrete wavelet transform. The leftmost image is the original image. The right image is the transformed one. The components are: top left compressed image C_t, top right horizontal coefficient image H_t, bottom left vertical coefficient image V_t and bottom right diagonal coefficient image D_t.

The energy is evaluated block wise dividing the image in regular blocks of fixed size and summing up the squared contribution coming from each coefficient image:

$$E(b_k, I_t) = \sum_{i,j \in b_k} V_t^2(i,j) + H_t^2(i,j) + D_t^2(i,j) \qquad (1)$$

where b_k is the k^{th} block in the input image I^t.

The energy value of a specific block varies significantly over time in presence or absence of smoke.

When the smoke covers part of the scene the edges are smoothed and the energy consequently lowered. This energy drop can be further emphasized computing the ratio $r(B_k)$ between the image energy of the current input frame and the one of the background model. The energy ratio has the advantage of normalizing the energy values and allowing a fair comparison between different scenes where the block energy itself can vary significantly. The ratio of the block b_k is given by:

$$r(b_k, I_t, Bg_t) = \frac{E(b_k, Bg_t)}{E(b_k, I_t)} \qquad (2)$$

where Bg_t is the background model up to time t and I_t is the input frame (see Fig. 4).

3.2 Color analysis to detect blended smoke regions

When a smoke event occurs, scene regions covered by smoke change their color properties. The smoke can either be completely opaque or partially transparent. In the former case the covered region changes completely its color while in the latter case the color of the covered region appears to be blended with the smoke color.

This simple observation remains valid in all the observed cases and intuitively suggests a hint to characterize the color of a smoke region.
The proposed model simply adopts an evaluation based on a blending function borrowed from computer graphics. A reference color model is chosen in the RGB color space to represent the color of the smoke in the scene. The model is selected by analyzing the different color tones produced burning different materials. For explanatory purposes is possible to concentrate the analysis to the case of a light gray color model as the smoke in the leftmost image of Fig. 3. Each pixel $I_t(i,j)$ of the input frame at time t is then checked against the smoke model and the background model Bg_t to evaluate the reference color presence computing the blending parameter bl using equation 3. The evaluation takes into account the case where the scene color and the smoke color are mixed together.

$$bl(i,j,I_t,Bg_t,S) = \frac{I_t(i,j) - Bg_t(i,j)}{S - Bg_t(i,j)} \qquad (3)$$

where Bg_t is the current background model at time t and S is the smoke reference color model.
To filter out the errors and possible measurements inaccuracy the blending value is computed for each image block as the average of bl values in the block:

$$\beta_{b_k}(I_t, Bg_t, S) = \frac{1}{N^2} \sum_{i,j \in \ ,b_k} \frac{I_t(i,j) - Bg_t(i,j)}{S - Bg_t(i,j)} \qquad (4)$$

where block size is $N \times N$
In conclusion the β measure quantifies how much each block globally shares chromatic properties with the reference color model.

3.3 A Bayesian approach for classification

In the previous subsections the block wise energy ratio measure r and the color blending measure β have been presented as possible discriminant features to identify a smoke region in the scene. A Bayesian formulation has been chosen to identify whether a block b_k is likely to belong to a smoke region. For each block the posterior probability of smoke presence, the event $f = 1$, considering the block b_k is defined:

$$P(f = 1|b_k) \propto P(b_k|f = 1)P(f = 1) \qquad (5)$$

The likelihood value is obtained by combining both the contributions coming from energy ratio and color information.

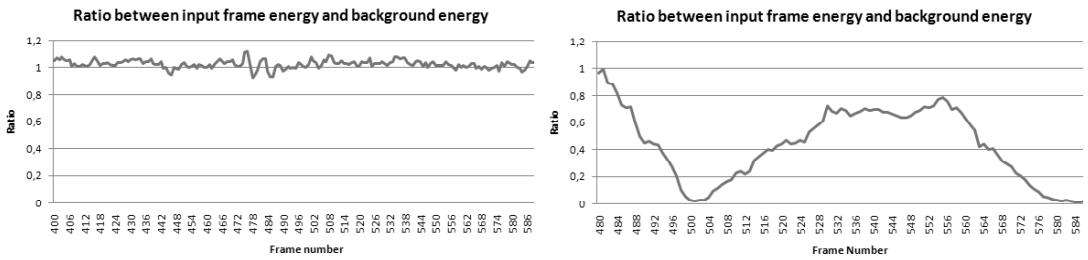

Figure 4: Left figure: the energy ratio trend of a non smoke block. Right figure: the energy ratio trend of a smoke block. In presence of smoke the energy ratio is subjected to gradual drops in its value.

These terms are considered probabilistically independent to simplify the treatment.

$$P(b_k|f=1) = P(r_{b_k}, \beta_{b_k}|f=1) = P_r(b_k|f=1) \cdot P_\beta(b_k|f=1) \quad (6)$$

The likelihood contribution due to energy ratio decay is obtained by summing the weighted Gaussians of the MOG having mean value below a considered threshold th_1 computed empirically observing the mean energy ratio value in smoke regions.

$$P_r(b_k|f=1) = \sum_{i=1}^{K} w_i N(r(b_k, I_t, Bg_t)|\mu_i\, \sigma_i) \quad (7)$$

when the i^{th} Gaussian mean value $\mu_i < th_1$.
The color contribution to the likelihood value is directly computed as the block color blending measure β_{b_k} according to equation 4.

$$P_\beta(b_k|f=1) = B_k(I_t, Bg_t, S) \quad (8)$$

The classification is biased making use of prior knowledge acquired observing several clips containing smoke. The prior probability of a smoke event in the current frame is directly related to the mean energy ratio value of the scene.

$$P(f=1) = \widehat{f}(\frac{1}{M} \sum_{\forall b_k \in I_t} r(b_k, I_t, Bg_t)) \quad (9)$$

where I_t is composed by M blocks.
The posterior probability value is thresholded to identify a candidate smoke block. The test for smoke presence is performed after foreground object segmentation. For any segmented object in the scene the number of candidate blocks intersecting the object's blob is computed. Finally an object is classified as smoke when the 70% of its area overlays candidated smoke blocks.

3.4 Test bed for experimental results

The proposed smoke detection system can be used in conjunction with a whichever video surveillance system providing moving object segmentation using a background model. The background model should be updated regularly but smoke regions should not be included in the background. This can be achieved choosing a slow background update rate and avoiding updating the background model areas where a smoke object is detected. The tests were performed using both the Stauffer and Grimson background model with selective update [13] and the SAKBOT median background model with knowledge based update [2]. Although the results did not vary significantly changing the background model and object detection technique, the second method has been preferred since discriminates the presence of possible shadows objects too.

			Temporal Analysis		Color Analysis		Gloabal Analysis	
Movie Name	Frame Number	Outdoor/Indoor	Time To Detect	False positive	Time To Detect	False positive	Time To Detect	False positive
movie_01	165	Outdoor	22	-	1	-	1	-
movie_02	210	Indoor	18	-	1	-	1	-
movie_03	2200	Outdoor	28	-	34	-	20	-
movie_04	3005	Indoor	212	-	273	-	185	-
movie_05	1835	Indoor	87	-	100	3	52	-
movie_06	2345	Outdoor	129	-	161	-	116	-
movie_07	2024	Indoor	57	3	99	-	35	-
movie_08	2151	Outdoor	88	2	88	-	42	-
movie_09	1880	Outdoor	59	-	56	-	45	-
movie_10	2953	Outdoor	457	-	498	-	300	-
movie_11	1485	Indoor	62	-	x	5	62	-
movie_12	499	Outdoor	43	-	8	-	16	-
movie_13	195	Indoor	53	-	23	-	27	-
movie_14	1226	Outdoor	77	-	370	-	69	-
movie_15	109	Outdoor	29	-	x	1	3	-

Figure 5: Experimental results of the proposed system on reference clips.

The system was tested on 50 clips of varying length in both indoor and outdoor setups where moving objects such as people or vehicles were present in the scene during the smoke event. Each clip contained a smoke event. Each likelihood term was evaluated separately to measure the impact on the system performance. The table in Fig. 5 summarizes the results obtained on 15 reference clips. The dataset is publicly available in the ViSOR system [14]. The first column of the table reports the video type and its frame-length. The average clips frame-rate is 25fps. The remaining columns report the results obtained using each likelihood term separately and finally the results of the whole system. The detection time (in frames) after the smoke event occurs is reported for all the test clips. The table clearly shows that the likelihood term due to temporal analysis (Eq.7) is effective in most of the observed cases. The main problem is the long detection time. This is caused by the time based statistics used to capture the energy ratio decay.

Although the likelihood contribution due to color blending has the advantage of speed up the detection process it tends to detect much false positives if used alone. See seventh column of Fig. 5. Observing the last two columns of Fig. 5 we can state that the complete approach is fast and reliable enough even in situations where each likelihood contribution fails. The overall system results on the 50 clips used for testing purposes report a detection rate of 77% 3 seconds later the smoke event occurs, 98.5% 6 seconds later and finally 100% 10 seconds later with an average true positive rate of 4%. Fig. 6 shows some snapshots of the system working on different conditions.

Figure 6: Snapshots of the proposed system working on several clips in different conditions. The blue area in the images is detected as smoke.

4. VISOR WEB INTERFACE

The ViSOR web interface has been designed in order to share the videos and the annotation contents. Some screen shots of the web interface are shown in Fig. 7. ViSOR supports multiple video formats, search by keywords, by video meta-data (e.g., author, creation date, ...), by camera information and parameters (e.g., camera type, motion, IR, omni-directional, calibration). Until now about 60 videos belonging to different scenarios, like indoor, outdoor, have been added to ViSOR but the number of video is growing day by day (In Table 4 some statistics about the ViSOR system are reported). Among them, 14 annotated videos of smoke detection have been added. Three modalities have been implemented to allow video access: video preview, based on a compressed stream, single screen shot (a representative frame of the entire video) or a summary view, in which clip level screen shots are reported. Two screen shots of the video representative of the smoke category are shown in Fig. 9 and Fig. 10 obtained using the Video and the Clip level mode respectively.

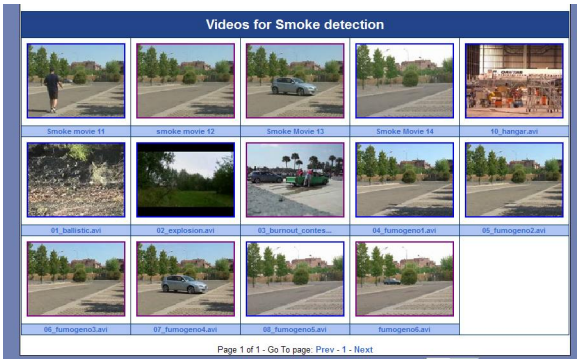

Figure 9: Thumbnails of the ViSOR videos belonging to the Smoke category using the Video Mode

Since annotations can be done at different levels of detail (smoke event detection only, events plus bounding box of the smoke, smoke and people annotation) and both ground truth and automatic annotations can be provided, for each video a set of annotations are shared and available for download. For each annotation, the entire annotation as well as a subset of the annotation fields, filtering by frame number, descriptor or single attribute can be extracted. At the present the annotation can be exported in the VIPER format only, but an MPEG7 format export module is under development.

Moreover, an integrated player for videos and annotations has been done. The player draws geometrical concepts superimposing them to the video; other linguistic concepts are reported below the video. A tree representation of the complete annotation content is reported as well. Finally, a set of descriptor level selection buttons are depicted in order to hide or show the relative annotation data. A screen shot of this player is reported in Fig. 8

Figure 10: Thumbnails of the ViSOR videos belonging to the Smoke category using the Clip Mode

Videos	
Uploaded Sequences	82
Annotated Sequences	31
Number of Clips	510
Concepts	
Videosurveillance IS-A concepts	96
IS-A concepts used in ViSOR	60
HAS-A concepts	36
Counters	
Web Accesses	80920
Video Downloads	1623

Table 4: Statistics

Figure 7: Some screenshots of the ViSOR web interface

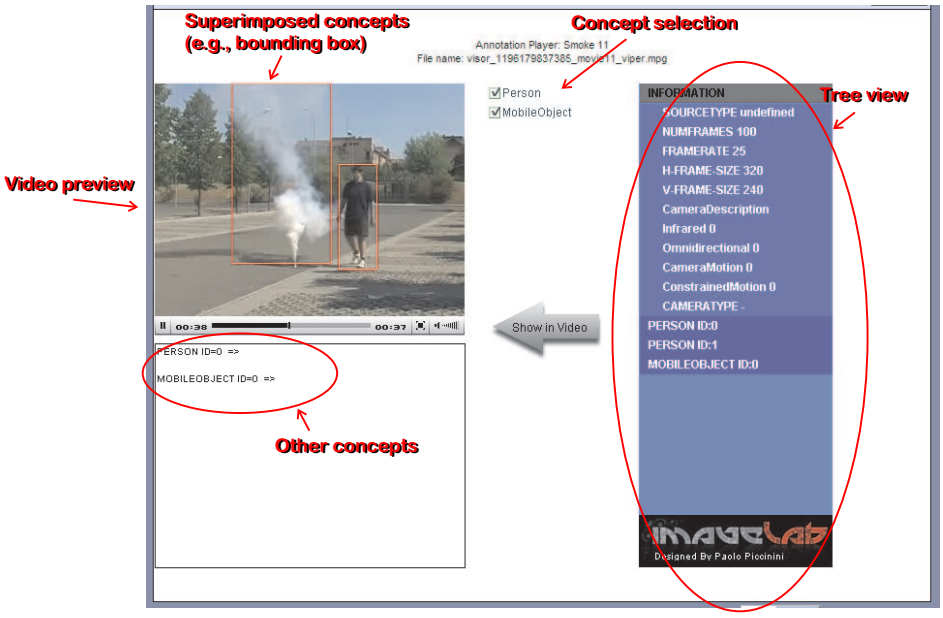

Figure 8: A screen shot of the flash player to preview video and annotation together

5. CONCLUSION AND FUTURE WORK

ViSOR is a dynamic 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 flexible structure and implementation of the system allows the exploitation on different application. For example, a section of the ViSOR system has been created to contains videos and annotation about smoke detection. In the paper the smoke detector framework used to generate the annotations is also described.

This project (funded by VidiVideo EU project) is recently started and even if the interface and the database structure have been developed, the population of the database is just on an initial stage. Nonetheless, its interactive interface and the free available tool set are key points to become a reference repository of surveillance and security videos for many multimedia applications.

6. REFERENCES

[1] H. O. S. D. Branch. i-lids - imagery library for intelligent detection systems. Website, 2006. `http://scienceandresearch.homeoffice.gov.uk/hosdb/`.

[2] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, Oct. 2003.

[3] D. Doermann and D. Mihalcik. Tools and techniques for video performance evaluation. *Proc. of Int'l Conference on Pattern Recognition*, 04:4167, 2000.

[4] A. R. Francois, R. Nevatia, J. Hobbs, and R. C. Bolles. Verl: An ontology framework for representing and annotating video events. *IEEE MultiMedia*, 12(4):76–86, 2005.

[5] L. Kennedy. Revision of lscom event/activity annotations, dto challenge workshop on large scale concept ontology for multimedia. Technical report, Columbia University ADVENT, 2006.

[6] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[7] M. Naphade, L. Kennedy, J. R. Kender, S.-F. Chang, S. J. R., P. Over, and A. Hauptmann. A light scale concept ontology for multimedia understanding for trecvid 2005. Technical report, IBM Research, 2005.

[8] R. Nevatia, J. Hobbs, and B. Bolles. An ontology for video event representation. In *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 7*, page 119, Washington, DC, USA, 2004. IEEE Computer Society.

[9] A.-T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin. Etiseo, performance evaluation for video surveillance systems. In *Proceedings of AVSS 2007*, 2007.

[10] I. Petrás, C. Beleznai, Y. Dedeoğlu, M. Pardàs, L. Kovács, Z. Szlávik, L. Havasi, T. Szirányi, B. U. Töreyin, U. Güdükbay, A. E. Çetin, and C. Canton-Ferrer. Flexible test-bed for unusual behavior detection. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 105–108. ACM, 2007.

[11] P. Piccinini, S. Calderara, and R. Cucchiara. Reliable smoke detection system in the domains of image energy and color. In *in press on 6th International Conference on Computer Vision Systems, Vision for Cognitive Systems*, 2008.

[12] C. Snoek, M. Worring, J. Van Gemert, J. Geusebroek, and A. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proceedings of the 14th ACM Int'l Conference on Multimedia*, pages 421–430, New York, NY, USA, 2006. ACM.

[13] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 246–252, 1999.

[14] Visor portal. Website, 2007. `http://imagelab.ing.unimore.it/visor`.

[15] Viper toolkit at sourceforge. Website, 2005. `http://viper-toolkit.sourceforge.net/`.

[16] *VSSN '06: Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, New York, NY, USA, 2006. ACM. General Chair-Jake K. Aggarwal and General Chair-Rita Cucchiara and Program Chair-Andrea Prati.

[17] Pets: Performance evaluation of tracking and surveillance. Website, 2000–2007. `http://www.cvg.cs.rdg.ac.uk/slides/pets.html`.