# Events detection using a video-surveillance Ontology and a rule-based approach

Yassine Kazi Tani[1], Adel Lablack[2], Abdelghani Ghomari[1], and Ioan Marius Bilasco[2]

[1] RIIR Laboratory, University of Es-Sénia, Oran, Algeria
yassine.kazi@gmail.com , ghomari65@yahoo.fr
[2] Laboratoire d'Informatique Fondamentale de Lille, Université de Lille 1, France
{adel.lablack, marius.bilasco}@lifl.fr

**Abstract.** In this paper, we propose the use of a Video-surveillance Ontology and a rule-based approach to detect an event. The scene is described using the concepts presented in the ontology. Then, the blobs are extracted from the video stream and are represented using the bounding boxes that enclose them. Finally, a set of rules have been proposed and have been applied to videos selected from PETS 2012 challenge that contain multiple objects events (e.g. Group walking, Group splitting, etc.).

**Keywords:** Ontology, Video surveillance, blobs, rules

## 1 Introduction

Nowadays, a growing amount of videos are available. This large amount of data that needs to be stored and indexed should be processed using efficient content based methods. Some of the existing works in video indexing use low-level features like color or motion for indexing video clips [5, 7]. Other approaches have their indexing system based on high-level features such as human interpretation using meta-data and keywords [15, 20]. These latter systems suffer from the exhaustive manual operations, and the semantic inconsistencies caused by different subjective interpretations made by people.

The semantic gap that exists between the low-level and the high-level features for an event could be solved by combining both levels using an ontology [8]. The use of ontologies for prior knowledge representation and scene understanding of video data is popular in many applications [12, 21, 22]. Gruber [9] defines the ontology as the representation of the semantic terms and their relationships. It consists of the representation of the concepts, their properties, and the relationship between concepts expressed in linguistic terms. The most important property is the derivation of an implicit knowledge through automated inference. It provides a formal framework to define domain knowledge [2].

We propose to use the concepts of a video surveillance ontology to derive rules that allows events detection from video sequences. The Ontology Web Language (OWL) [17] has been used to represent our ontology and the Semantic Web Rule Language (SWRL) [13] to generate the inference rules.

The remainder of this paper is organized as follows. Section 2 reviews some related work in the field of video processing using ontologies. In Section 3, we describe the architecture of our ontology of the video surveillance domain. We describe the methodology used to derive the rules based on the video surveillance domain ontology in Section 4 using the PETS 2012 dataset as a case of study. Finally, we give concluding remarks and potential future work in Section 5.

## 2 Related work and background

Several works based on an ontology have been proposed to overcome the semantic gap between low-level and high-level features. Bagdanov et al. [1] present a system to solve the semantic gap between the high-level concepts and the low-level descriptors using a multimedia ontology. It contains visual prototypes that represent each cluster and act as a bridge between the domain ontology and the video structure ontology. Dasiopoulou et al. [8] have used color homogeneity as descriptor. The visual objects have been included in the ontology and the semantic concepts have been derived from color clustering and reasoning. Bertini et al. [3] have used both generic and domain specific descriptors to identify visual prototypes that represent elements of visual concepts. New instances of visual concepts are then added to the ontology through an updating mechanism of the existing concepts. Finally, the prototypes are used to classify the events and the objects that are observed in video sequences.

In video surveillance applications, some specific events like abnormal events have to be detected from streams provided generally by stationary cameras. An ontology can be used to support the indexing process. Xue et al. [21] proposed an ontology-based surveillance video archive and retrieval system. Lee et al. [10] implement a framework called Video Ontology System (VOS) to classify and index video surveillance streams. Snidaro et al. [18] have used a set of rules in SWRL language for event detection in video surveillance domain. In order to overcome the problem of the manual rules creation by human experts, Bertini et al. [4] proposed an adaptation of the First Order Inductive Learner technique (FOIL) for Semantic Web Rule Language (SWRL) named FOILS.

Most of the previous works in the surveillance domain have used the ontology tool and demonstrate its efficiency to help and manage the indexing and retrieval process. They consider events such as abandoned object, stolen object, a person who is walking from right to left, an airplane that is flying, etc. SanMiguel et al. [11] have proposed an ontology for representing the prior knowledge related to a video event analysis. It is composed of two types of knowledge related to the application domain and the analysis system. Domain knowledge involves all the high level semantic concepts (objects, events, context, etc.) while system knowledge involves the abilities of the analysis system (algorithms, reactions to events, etc.). However, this ontology determines only the best visual analysis framework (or processing scheme) without any inference reasoning for objects tracking and events detection or analysis.

In this paper, we propose to use an ontology based-approach to detect single/multiple objects events through a set of SWRL rules. It allows the transition from the blobs extracted using visual analysis module to the detection of an event.

## 3 The architecture of the Ontology

The ontology approach is an effective way to support various processes for events detection in video surveillance domain. The scene is described using the concepts presented in the ontology and a video analysis module extracts the blobs from the streams using some low level property such as color, position, size, etc. The ontology considers these blobs as an input through the bounding boxes that enclose them and instantiate their features for creating the different DataType Property in the ontology. Then, the reasoner of our ontology classifies, in the first step, the different bounding boxes in their respective semantic meaning (Group_Of_Person/ Person) using a set of SWRL rules [13] and associates, in a second step, this video sequence, using another set of SWRL rules, to the appropriate video event class regarding the behavior of its objects.

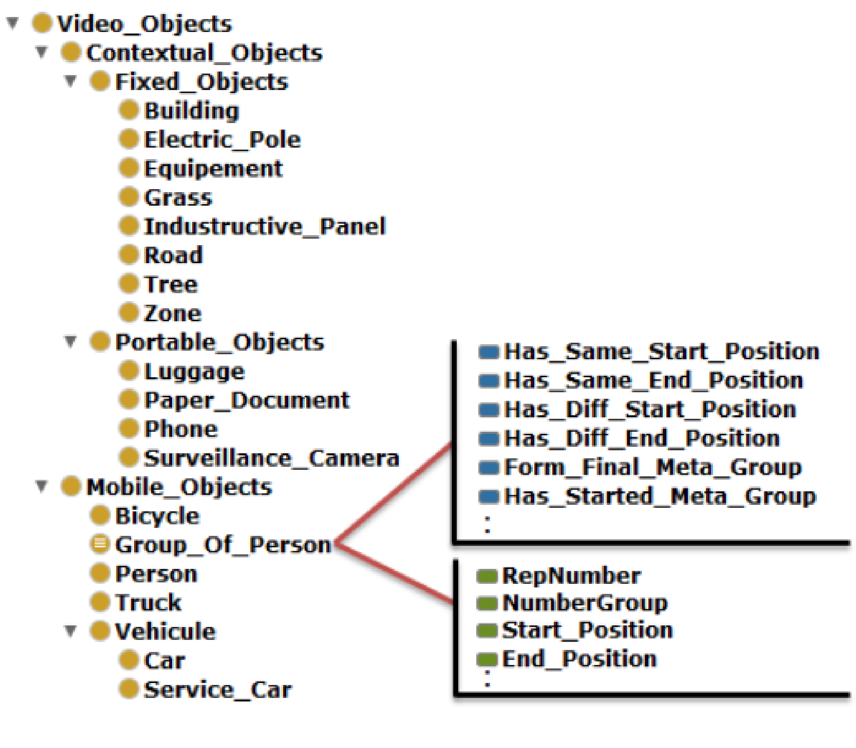


**Fig. 1.** Video_Objects class hierarchy sample.

In order to have an efficient representation of the video surveillance domain, we preserved the same organization proposed by SanMiguel et al. [11] and complete it by adding new concepts. We organize our ontology in four categories, ranging from high-level concepts to low-level features : Video Events (gather all events that can happen in the video surveillance domain), Video Objects (represents a set of objects that can appear in a video sequence), Video Sequences (all the video sequences that could be indexed by our Ontology) and Bounding boxes (all the bounding boxes that enclose the blobs detected by the video analysis module in a video sequence with their low level features). The Figure 1 depicts a sample of the Video_Objects class hierarchy.

# 4 The rule based approach

In this section, we propose to use the PETS 2012 dataset as a case of study to depicts our rule based approach that allows to handle a video surveillance ontology for events detection in video streams.

## 4.1 PETS 2012 dataset

A set of events selected from PETS 2012 challenge [6] are used to experiment the efficiency of the proposed rules. This dataset contains different crowd activities and the task is to provide a probabilistic estimation of some events and to identify the start and the end of the events as well as transitions between them.

## 4.2 Scene representation

In order to determine the best configuration of the processing schemes for detecting the events, we describe the scene in terms of concepts of our ontology. The Figure 2 shows an ideal and very precise segmentation of two scenes extracted from PETS 2012 challenge. Although some automatic techniques might be use for segmentation, we have started from a manual segmentation of the scene as the scene contains static elements that will not change over time (building, grass, electric pole, road, trees, car parks, restrictive roads). These elements have a strong semantic meaning, that can enhance the reasoning process and interpret the events resulting from other (volatile) elements (service car) that are subject to movements within the scene setting. For instance, special attention should be raised if moving objects are present in the Restrictive Road and deep analysis should be run to see if the moving objects are pedestrian or cars. Changes in appearances of studied objects can also be relevant in extracting meaningful events (a tree going reddish, might be a strong feature in detecting an abnormal event). Although, we are more focusing on movement reasoning, both kinds of changes (movement and appearance) result in the presence of regions yielding similar characteristics in terms of appearance and/or motion commonly called blobs.


**Fig. 2.** Scene Representation from PETS 2012 challenge camera views.

### 4.3 Blobs extraction

We propose an event detection approach based on blob regions. Blobs have proven to be a better feature cue than points, corners or edges as they usually have a larger coverage area and total occlusion of the subject is more unlikely to happen. So, we should identify all the major blobs in the scene. A major blob is defined as a blob that shows potential area size to be considered [19, 16]. This is an essential step towards determining potential person/group.

In order to collect these blobs, several algorithms could be used. A background subtraction algorithm will classify the pixels of the input image into foreground and background. Then the blobs are extracted by groping together the foreground pixels belonging to a single connected component. We can also use optical flow by extracting the characteristics of each pixel in each motion image. These flows are then grouped into blobs that have coherent motion and are modeled by a mixture of multivariate Gaussians. The optical flow is useful to characterize each moving pixel according to certain features of the flow vector.

The Figure 3 highlights the bounding boxes that enclose the detected blobs in different situations like Group walking, Group running, Group Splitting, etc.

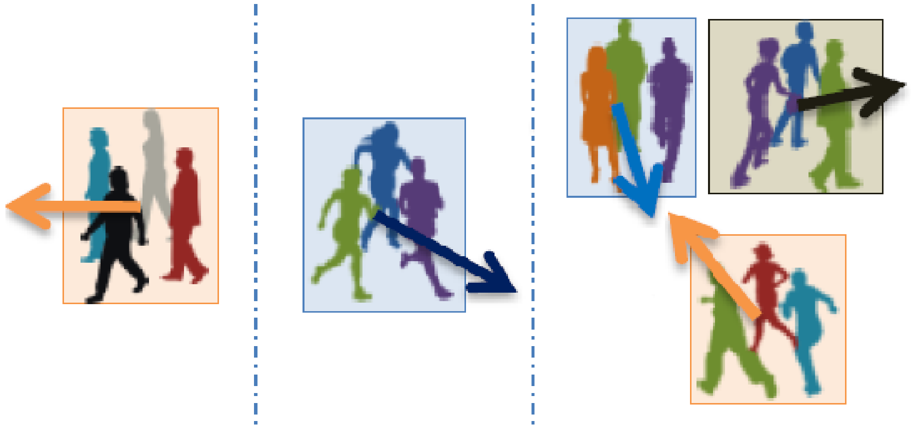

**Fig. 3.** Events from PETS 2012 challenge: Group walking, Group running, Group merging and Group splitting.

A pre-processing stage is often applied to select the major blobs. It is done by applying some anthropomorphic assumptions and morphological operations. The following morphological operations are performed:

- Closing: Morphological closing smoothness sections of contours, fuse together narrow breaks and long gulfs.
- Fill holes: A flood-fill operation is performed to close up the remaining small holes.
- Removal of motion at boundary: Pixels of the motion region that are located along the boundary are eliminated to avoid ambiguity of the region belonging to a possible moving object.

At this stage each blob can represent either an entire object, an object subpart or can be generated by noise. It is identified by a label and the surrounding bounding box. These bounding boxes are then used as input for the rule stage. The aim of this rules is to ensure the identification of semantically significant objects by analysing detected blobs over consecutives frames.

By comparing the bounding boxes found in two consecutive frames, our rule based approach is able to assess for each blob of the previous frame if it has been found or if undergoes a split or takes part in a merger. It consists in establishing the associations between the objects found in the previous frame and the blobs just extracted and grouped within the bounding boxes. We describe now our strategy according to the blobs that have been detected in the current frame:

- Straightforward tracking: this is the simplest case and it corresponds to two blobs without neighboring ones which are detected approximately in the same position in two successive frames and there are no splits nor merges (blob size is preserved or slightly varies). The concept of approximately in the same position is implemented trough the definition of a threshold on a distance measurement between the blobs.

- Splitting: a split is detected when a blob breaks in two distinct ones. We validate every split as soon as it occurs, creating two new objects. However, the original object identity is resumed if this fragmentation of the object into two blobs is temporary which may be due, for example, to an error during the detection phase.

- Merging: we detect a merging event when two objects having close past trajectories and detected up to frame at time t-1 merge their bounding boxes in the frame at time t. If these conditions are satisfied, the algorithm creates a new object joining the trajectories of the two previous ones

Some events that could happen may introduce a confusion in this process such as:

- Disappearance: an object detected in a frame at time t-1 is classified as lost in the current frame if no blob is present in the neighbourhood of the expected object position at time t. If an object is lost in proximity of an image border, the algorithm assumes that the object has left the scene, else waits for the appearance of the object in proximity of the place where it disappeared. Still, we should ensure that no other blob belonging to another semantically significant object was/is around, and takes the place of the previous.

- Occlusion: it is distinguished from merging/splitting events on the basis of the direction of the past trajectories. When an occlusion occurs, we wait to analyze the scene for a specific number of frames to find the correct association between the objects found before and after the occlusion.

### 4.4 The rules construction

Different events from the PETS 2012 challenge could be used to depict the efficiency of the proposed approach such as:

- Group running and walking events: it consists to estimate if the people forming a group are walking or running. These events can be identified using the motion magnitude in each image. High magnitude event means running while a low magnitude means walking event. The detection is done either by defining an experimental threshold or using a classifier with feature such as the average speed of movement.
- Group formation and splitting events: it consists in the detection and the analysis of the position, the orientation and the speed of the groups.

We have used the Rule plugin of Protégé [14] to write the inference rules of our engine in SWRL language. Our rules are divided into 3 categories:

- Distance rules: it consists on checking the distance between the detected bounding boxes in the current frame. The bounding boxes that are close to each other are grouped into a major bounding box.
- Tracking rules: it consists on tracking the major bounding boxes generated by the previous category over the frames to detect the start/end position.
- Event rules: it consists in analyzing the behaviour of the groups identified in the previous category in order to detect the appropriate event.

The left side of the rule (before the arrow) is checked by the inference engine and the reasoner infer or not the right side. The Figure 4 depicts the construction of a distance rule. It checks if two bounding boxes could be grouped into a major bounding box.


BB(?BBx), BB(?BBy), Frame(?F1), MBB(?MBB1), MBB(?MBB2), BB_Detected_In_Frame(?BBx, ?F1), BB_Detected_In_Frame(?BBy, ?F1), BB_Bottom_Left_Point_Y(?BBx, ?h), BB_Bottom_Right_Point_Y(?BBy, ?d), BB_Number(?BBx, ?n4), BB_Number(?BBy, ?n5), BB_Top_Left_Point_X(?BBx, ?a), BB_Top_Left_Point_X(?BBy, ?f), BB_Top_Left_Point_Y(?BBx, ?e), BB_Top_Left_Point_Y(?BBy, ?l), BB_Top_Right_Point_X(?BBx, ?j), BB_Top_Right_Point_X(?BBy, ?b), BB_Top_Right_Point_Y(?BBy, ?c), MBB_ID(?MBB1, ?n1), MBB_ID(?MBB2, ?n1), Number_BB_In_Frame(?F1, 2), Number_Frame(?F1, ?n1), Number_MBB(?MBB1, ?n2), Number_MBB(?MBB2, ?n3), add(?x2, ?b, 20), greaterThan(?a, ?b), greaterThan(?h, ?d), greaterThan(?n3, ?n2), greaterThanOrEqual(?b, ?x1), greaterThanOrEqual(?e, ?c), lessThanOrEqual(?a, ?x2), lessThanOrEqual(?e, ?d), subtract(?x1, ?a, 20), subtract(?z1, ?j, ?f), subtract(?z2, ?h, ?l) -> BB_Represent_MBB(?BBx, ?MBB1), BB_Represent_MBB(?BBy, ?MBB1), MBB_Detected_In_Frame(?MBB1, ?F1), MBB_H(?MBB1, ?z1), MBB_Top_Left_Point_X(?MBB1, ?f), MBB_Top_Left_Point_Y(?MBB1, ?l), MBB_W(?MBB1, ?z2)

**Fig. 4.** A rule for grouping two bounding boxes into a major bounding box.

This rule presented above is constructed as follow: (i) The reasonner checks in the current frame if the positions of the two bounding box ($BB1$, $BB2$) are close in the $X$ and $Y$ axis. The Bounding boxes are then tested as $BBx \rightarrow BB2$ and as $BBy \rightarrow BB1$ using the following conditions: (i) $Top\_Right\_Point\_Y BB1 \leqslant Top\_Left\_Point\_Y of BB2 \leqslant Bottom\_Right\_Point\_Y of BB1$, (ii) $Top\_Left\_Point\_X of BB2 \leqslant Top\_Right\_Point\_X of BB1 + 20$ and $Top\_Right\_Point\_X of BB1 \geqslant Top\_Left\_Point\_X of BB2 + 20$. In this case, the reasoner will infer that both bounding boxes belong to the same Major Bounding Box and updated it.

A large set of rules is proposed to model all the situations that could happen in the scene according to the events handled by our ontology. The output of each category could be used as input for another one. Indeed, an event is detected using a rule that took as input the information inferred by a tracking rule that has been applied to major bounding boxes identified using distance rules.

The inherent difficulty of writing down rules in SWRL or equivalent language is the fact that the events are spanning over various time intervals. Various time windows can be applied to the same event detection. A split event can occur in a very short time-frame, if the groups are evolving at high speed or it could take a long time-frame if the groups are evolving at low speed. However, we are using a fixed time-window in order to simplify writing rules.

## 5 Conclusion

Video Surveillance systems become popular in our daily life to ensure security and safety and allows to study human behavior. In this paper, we have presented our rule based approach that allows to handle a video surveillance ontology to detect single or multiple objects events.

In our future work, we will extend our ontology to model new concepts and improve our SWRL rules for handling different events that can occur in video surveillance domain.

## References

1. Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo, Giuseppe Serra, and Carlo Torniai. Semantic annotation and retrieval of video events using multimedia ontologies. In *International Conference on Semantic Computing (ICSC)*, pages 713–720, 2007.
2. Lamberto Ballan, Marco Bertini, Alberto Del Bimbo, and Giuseppe Serra. Semantic annotation of soccer videos byvisual instance clustering and spatial/temporal reasoning in ontologies. *Multimedia Tools and Applications*, (2):313–337, June 2010.
3. Marco Bertini, Alberto Del Bimbo, Costantino Grana Carlo Torniai, and Rita Cucchiara. Dynamic pictorial ontologies for video digital libraries annotation. In *1st ACM Workshop on The Many Faces of Multimedia Semantics*, pages 47–56, 2007.
4. Marco Bertini, Alberto Del Bimbo, and Giuseppe Serra. Learning ontology rules for semantic video annotation. In *2nd ACM Workshop on Multimedia Semantics*, 2008.
5. Alberto Del Bimbo, Pietro Pala, and Enrico Vicario. Spatial arrangement of color flows for video retrieval. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 413–416, 2001.
6. PETS 2012 challenge. [online]: http://www.cvg.rdg.ac.uk/pets2012/a.html, 2012.
7. Bertrand Chupeau and Ronan Forest. An evaluation of the effectiveness of color attributes for video indexing. In *SPIE Storage and Retrieval for Media Databases*, pages 470–481, 2001.

8. Stamatia Dasiopoulou, Vasileios Mezaris, Ioannis Kompatsiaris, V.-K. Papastathis, and Michael G. Strintzis. Knowledge assisted semantic video object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, (10):1210–1224, Oct 2005.
9. Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, (5-6):907–928, Nov-Dec 1995.
10. Jeongkyu Lee, Munther H. Abualkibash, and Padmini K. Ramalingam. Ontology-based shot indexing for videosurveillance system. In *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering*, pages 237–242, 2008.
11. Juan Carlos San Miguel, José María Martínez Sanchez, and Alvaro García-Martín. An ontology for event detection and its application in surveillance video. In *6th IEEE International Conference Advanced Video and Signal based Surveillance (AVSS)*, pages 220–225, 2009.
12. Natalya F. Noyet and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical report, 2001.
13. Martin O'Connor, Holger Knublauch, Samson Tu, Benjamin Grosof, Mike Dean, William Grosso, and Mark Musen. Supporting rule system interoperability on the semantic web with swrl. In *4th International Semantic Web Conference (ISWC)*, pages 974–986, 2005.
14. Protégé. The protégé project, [online]: http://protege.stanford.edu, 2012.
15. Juan M. Sanchez, Xavier Binefa, Jordi Vitria, and Petia Radeva. Linking visual cues and semantic terms under specific digital video domains. *Journal of Visual Languages and Computing*, 11(3):253–271, 2000.
16. John See, Lee Sze Wei, and Madasu Hanmandlu. Human motion detection using fuzzy rule-base classification of moving blob regions. In *International Conference on Robotics, Vision, Information and Signal Processing (ROVISP)*, 2005.
17. M. K. Smith, C. Welty, and D. McGuinness. Owl web ontology language guide. In *W3C Recommendation , http://www.w3.org/TR/2004/REC-owl-guide-20040210/*, 2004.
18. Lauro Snidaro, Massimo Belluz, and Gian Luca Foresti. Representing and recognizing complex events in surveillance applications. In *4th IEEE International Conference Advanced Video and Signal based Surveillance (AVSS)*, pages 493–498, 2007.
19. Luigi Di Stefano, Martino Mola, Giovanni Neri, and Enrico Varani. A rule-based tracking system for video surveillance applications. In *International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES)*, 2002.
20. Yi Wu, Yueting Zhuang, and Yunhe Pan. Content-based video retrieval integrating human perception. In *SPIE Storage and Retrieval for Media Databases*, pages 562–570, 2001.
21. Ming Xue, Shibao Zheng, and Chongyang Zhang. Ontology-based surveillance video archive and retrieval system. In *5th International Conference on Advanced Computational Intelligence (ICACI)*, 2012.
22. Jawahir Che Mustapha Yusuf, Mazliham Mohd Su' ud, Patrice Boursier, and Muhammad Alam. Extensive overview of an ontology-based architecture for accessing multi-format information for disaster management. In *International Conference on Information Retrieval and Knowledge Management (CAMP)*, pages 294–299, 2012.