Towards a Generic Ontology for Video Surveillance

Pablo Hernandez-Leal, Hugo Jair Escalante(✉), and L. Enrique Sucar

Instituto Nacional de Astrofísica, Óptica y Electrónica Sta. María Tonantzintla, Puebla, Mexico
{pablohl,hugojair,esucar}@inaoep.mx

Abstract. Video surveillance is an important problem that has been studied for several years. Nowadays, in the context of smart cities, intelligent video surveillance is an important topic which has several subproblems which need to be solved and then integrated. For example, on one side there are several algorithms for detection, recognition and tracking of objects and people. On the other side, it is necessary to recognize not only objects and persons but complex behaviors (fights, thefts, attacks). To solve these challenges, the use of ontologies has been proposed as a tool to reduce this gap between low and high level information. In this work, we present the foundations of an ontology to be used in an intelligent video surveillance system.

Keywords: Video surveillance · Ontology · Event recognition

1 Introduction

Ontologies are used to represent knowledge and they have different uses. In video recognition and analysis, their main use has been to remove the *semantic gap*, this is, the difficulty of mapping semantic concepts into a set of image and/or spatiotemporal features that can be automatically extracted from video data without human intervention [2]. For example, even when a concept such as "car accident" is an accepted standard linguistic concept, there are thousands of videos that fulfill that linguistic term.

An ontology for analyzing this sort of videos would specify all associated entities and how they relate to each other. Ontologies are basically defined by two parts: entities and its relationships.

There are several works in video surveillance that use ontologies, one of them proposes a multi-camera platform for bank monitoring [3]. The platform follows the steps: (i) Object detection for each camera. (ii) Tracking for each camera to obtain a graph of moving objects. (iii) A fusion of information is used to combine the information from different cameras producing a global graph. (iv) This global graph is used to perform long term analysis (i.e., behavior recognition). (iv) Contextual information can be added (i.e., 3d information).

Previous works have used ontologies to detect objects and events. Some were used together with a rule detection system [7]. In that work, the used data was

E. Sucar et al. (Eds.): AFI 2016, LNICST 179, pp. 3–7, 2017.
DOI: 10.1007/978-3-319-49622-1_1

part of the PETS (Performance Evaluation of Tracking and Surveillance) 2012 competition where the objective was to estimate the probability of certain events as well as its initial and final times.

Ontologies for specific applications had been designed, for example for analyzing soccer videos [1]. In this ontology each concept is associated with a visual concept. Concepts are clustered by their spatiotemporal similarity and each cluster has a visual prototype (clip, shot, frame or part of a frame).

An approach to use an ontology for video surveillance was presented by San-Miguel et al. where the ontology is based on two levels of knowledge: scene (domain information) and system (analysis information) [6]. In the scene there are three types of entities: (i) events: can be simple or complex, (ii) context: spatial information and (iii) objects: movables and context. In the system part there are different categories: (i) status, (ii) capacity (e.g., input parameters, output) and (iii) reaction (i.e., record event, activate alarm). The authors evaluated the approach in recognizing events such as grabbing and dropping objects.

In this work we contribute with a generic ontology for video analysis, we present the main components and relationships that are needed in a video surveillance environment.

2 Proposed Ontology

The proposed ontology is expected to be used jointly with a surveillance system.

2.1 Ontology Use

In particular, we want to use the ontology to help in the detection of events, and behaviors. An ontology is useful to define all the objects and behaviors that are important in video surveillance. The ontology can help to decide which algorithms to use [6]. Further, it will be interesting to use the ontology as a way to retrieve videos that match specific events [5]. In the context of smart cities there are another set of possible uses of an ontology in video analysis, namely: automatic annotation of videos, automatic extraction of videos and automatic annotation of composed events, among others [1]. Finally, another idea is to use the structure of the ontology together with hierarchical classifiers.

Based on the previous works and with the objective of developing a generic ontology for video surveillance we present our proposal. The ontology was defined using Protégé [4].

2.2 Entities

Entities are divided in three groups: Content, System and Context.

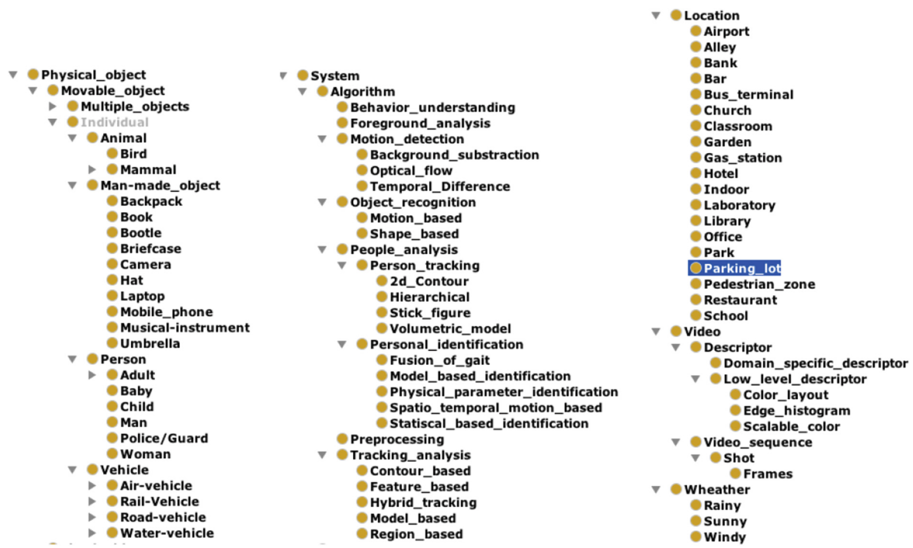
Content. There two important groups: `Event` and `Physical_object`. There are two types of `Event`: `Person_Event` and `Vehicle_Event`. Inside `Person_Event` there are activities such as: `Standing`, `Running`, `Sitting`, `Walking`, etc. There is also another group for generic events: such as `Enter` or `Exit` an area. Examples of `Person_event` are: `Greeting`, `Fighting` or `Hugging`. Inside `Physical_object` there are two groups: `Movable_object` and `Fixed_objects` (see Fig. 1 (a)).

There are two types of `Movable_object`: `Individual` and `Multiple_objects`. In the `Individual` category there are animals (`Bird`, `Mammal`), persons (`Baby`, `Adult`, `Man`, `Woman`), vehicles (`Air`, `Road`) and personal objects (`Backpack`, `Hat`, `Laptop`, `Camera`, `Book`). A `Multiple_object` is a group of persons, vehicles or animals.

Fixed objects are considered those that do not change naturally of place, for example: generic (`Building`, `Fence`, `Furniture`, `Window`, etc.), natural object (`Vegetation`, `Sky`, `Snow`, etc.) and road objects (`Sidewalk`, `Street`, `Traffic_light`, etc.).

System. This group contains information that the video surveillance system will use. There are 4 entities: 3 auxiliary concepts `Point`, `Reaction` and `Area`, and another entity `Algorithm`.

There are different sub-entities of `Algorithm` that describe different tasks that are needed in video surveillance: preprocessing, object detection, object recognition, tracking, behavior analysis, among others. Also inside each entity there can be another division depending on what the technique is based on (see Fig. 1 (b)).



(a) (b) (c)

Fig. 1. Parts of the proposed ontology: (a) `Object`, (b) `System` and (c) `Context`.

Context. Here is where the information regarding the environment is defined (see Fig. 1 (c)). For example weather (`Rainy`, `Sunny`, `Windy`), location (`School`, `Church`, `Parking`, `Bank`, etc.) and information from the video itself (`Frame`, `Shot`, `Descriptor`, etc.).

2.3 Properties

Besides entities, properties are important in an ontology. In particular, there two important types of properties: object properties and datatype properties.

Object Properties. These are defined between instances of entities. An example is `hasArea`. Entities such as `Event` or `Physical_object` have an `Area`. This is useful, for example, when the system detects an object and in the video we want to highlight a box to show that object, this is represented by the object `Area`. Another example is: `hasReaction`. Every `Algorithm` has a possible reaction such as activate an alarm, record the event, or do nothing.

Datatype Properties. The second type of properties relate entities and datatypes (`int`, `float`, `string`, etc.). An example of this type of property is `hasSpeed` which refers to a `Movable_object` that is moving and this speed is represented internally by a `float` value. Another example is the property `hasTitle` which belongs to every `Video` that has a title represented as a `string`.

3 Conclusions and Future Work

In this work we proposed an ontology for video surveillance. The idea is to reduce the semantic gap among linguistic and visual concepts. As future work we plan to combine the ontology with algorithms for detection and tracking to recognize important behaviors in video surveillance.

References

1. Bagdanov, A.D., Bertini, M., Bimbo, A.D., Serra, G., Torniai, C.: Semantic annotation and retrieval of video events using multimedia ontologies. In: International Conference on Semantic Computing, Irvine, California, USA, pp. 713–720, July 2007
2. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.K., Strintzis, M.G.: Knowledge-assisted semantic video object detection. IEEE Trans. Circuits Syst. Video Technol. **15**(10), 1210–1224 (2005)
3. Georis, B., Maziere, M., Bremond, F.: A video interpretation platform applied to bank agency monitoring. In: Intelligent Distributed Surveilliance Systems, London, UK, pp. 46–50 (2004)
4. Knublauch, H., Fergerson, R.W., Noy, N.F., Musen, M.A.: The protégé OWL plugin: an open development environment for semantic web applications. In: McIlraith, S.A., Plexousakis, D., Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 229–243. Springer, Heidelberg (2004). doi:10.1007/978-3-540-30475-3_17

5. Mezaris, V., Kompatsiaris, I., Boulgouris, N.V., Strintzis, M.G.: Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. IEEE Trans. Circuits Syst. Video Technol. **14**(5), 606–621 (2004)
6. SanMiguel, J.C., Martínez, J.M., García, A.: An ontology for event detection and its application in surveillance video. In: International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, pp. 220–225 (2009)
7. Kazi Tani, M.Y., Lablack, A., Ghomari, A., Bilasco, I.M.: Events detection using a video-surveillance ontology and a rule-based approach. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8926, pp. 299–308. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16181-5_21