An Unified Knowledge Representation Framework for Surveillance Videos

Krishna Chandramouli*, Virginia Fernandez Arguedas+, Ebroul Izquierdo+

*Division of Enterprise and Cloud Computing,
School of Information Technology and Engineering,
Vellore Institute of Technology University,
Vellore, India 632014.
email: krishna.c@vit.ac.in

+ Multimedia and Vision Research Group,
School of Electronic Engineering and Computer Science,
Queen Mary, University of London, Mile End Road, E1 4NS, London, UK.
email: {virginia.fernandez, ebroul.izquierdo}@eecs.qmul.ac.uk

Abstract

Knowledge Representation in the domain of Surveillance has attracted the attention of many researchers from interdisciplinary areas. Several fragmented ontologies are found in the literature for describing events and indexing media items to facilitate semantic retrieval and object annotation. However, with the ever increasing deployment of CCTV throughout the world, it is imperative that ontologies are developed that could unify the relevant and necessary information that are obtained from surveillance installation. Hence, in this paper we present a comprehensive knowledge representation framework for modelling, indexing, classifying and retrieving surveillance videos for forensic applications. The framework integrates multitude dimension of information including geospatial grid representation, temporal semantics, event representation and object annotation through multimedia ontology. The framework thus would enable querying of high-level information that otherwise would not be cross-indexed.

Keywords: Surveillance Ontology, Geospatial Object Tracking, Event Detection, Semantic Retrieval System, Knowledge Representation

1 Introduction

Recent technology development coupled together with peoples' concern for safety and security have caused a wide spread application of Closed Circuit Television (CCTV). These cameras provide a sense of security through its constant monitoring of the installed spatial locations. As a consequence of constant monitoring these cameras produce large amount of video data sets which, without further processing would consume a lot of manual labour to detect any anomaly experienced [3]. An example of a video footage being generated is shown in Fig. 1. In literature, there are several analysis algorithms reported for extracting semantics from such surveillance footage.

Figure 1. Surveillance Video of a CCTV

Video analysis module typically provides two main types of result: objects and events. Thus, surveillance video indexing and retrieval approaches can be divided into two categories: surveillance video indexing and retrieval at the object level [5], [9], [7] and at the event level [14], [13]. As events of interest may vary significantly among different applications and users, this chapter focuses on presenting the work done for surveillance video indexing and retrieval at the object level. In terms of algorithmic categorization, the following are the main research areas.

- Object/Person Recognition with Multi-view Tracking
- Event Recognition
- Semantics based Surveillance Retrieval System

In each and every one of the research field, there has been many developments of significance. In [12], authors present a computational approach towards abnormal visual event detection based on exploring and modeling local motion patterns in a non-linear space. Provided with a small sample of annotated normal motion vectors, the non-linear detector ranks segments in a sequence as a function of abnormality. In an another work [2], authors present an algorithm for the automated external calibration (localization) of a network of cameras with non-overlapping fields of view, a type of network that is becoming widespread for monitoring large environments. As technol-

ogy develops, researchers have focussed on combining different field of view information obtained from camera network. In such networks, object tracking is an important step in many applications related to security, traffic monitoring and event recognition. Such applications require the optimal trade-off between accuracy, communication and computing across the network. The costs associated to communication and computing depend on the type and amount of cooperation performed among cameras for information gathering, sharing and processing to validate decisions as well as to rectify (or to reduce) estimation errors and uncertainties. Addressing this problem, an extensive survey article is presented by [1].

Similarly, on the knowledge representation and semantics front, the Ontology design models have evolved to distinguish between things and happening. This distinction facilitates effective study of event-oriented geographic phenomena [8]. In [11], authors propose a space-time thematic Ontology which supports some aspects of event queries. In terms of semantic representation, [6] has categorised events into natural events, artificial events, etc. Although the concept of event is repeatedly used in the construction of the Ontology, in [3], authors used the concept as a central element from which they build the rest of the knowledge resources. The set of events is organised as the central taxonomy of the Ontology.

The mid-level information extracted from the analysis tools and components are stored in non-interchangeable format and hence could not cater to the needs of human queries. In this paper, we present an unified knowledge representation framework for surveillance videos that could cater to the questions posed by humans. From Fig. 1, in addition to the objects and events that could be extracted, there are additional complementary knowledge embedded in the video that provides spatio-temporal context for both the CCTV camera and the events occurring in the video. Some examples of the information encapsulated within the video are the spatial location of the camera, the physical correspondence of the camera view and the neighborhood relationship between such cameras. The unified knowledge representation framework, namely CCTV Ontology presented here facilitates formalization of both contextual, media and event related semantic metadata. In addition, the inclusion of time ontology enables synchronization of multiple components and forms a link between extracted knowledge informations.

The remainder of the paper is structured as follows. In Section 2, an overview of the different ontology models is presented. The structure of these ontologies are discussed beyond surveillance domain and modeling (or) mapping the ontology to represent surveillance specific information presents a challenge in itself. In Section 3, the proposed knowledge representation framework is presented. Finally, Section 4 presents the conclusion and future work of the paper.

2 Literature Review

In the literature, there are several ontology models proposed for representing information extracted from the analysis algorithms. In general, representation and recognition of events in a video is highly relevant for diverse research fields such as video surveillance, video browsing and content based video indexing. In order to achieve interoperability between the research community and the users of surveillance systems a common representation for describing events as it would then allow easy interchange of video annotations and sharing of video recognition modules among researchers. Advanced Research and Development Activity (ARDA) group members have developed of a formal language for describing an ontology of events, which referred as VERL (Video Event Representation Language) and a companion language called VEML (Video Event Markup Language) to annotate instances of the events described in VERL.

On the other hand, mapping spatial data requires the extensive use of map scale [10]. Such map oriented ontology representation exists in CityGML, where there are several properties connecting an object to its geometry corresponding to five different levels of detail important in city models (e.g. lod0Geometry, lod1Geometry, etc.). In a cartographic generalization process a very important property attached to a representation is that of minimal dimension and minimal distance so that a detail is legible and two features can be distinguished. A typology of constraints on a process is proposed by [4]: topology, position/orientation, shape, pattern, distribution/statistic. Some constraints explicitly refer to relations (topology, orientation) or properties (shape) and others do not but are expressed, in the computing model, as relations and properties. To summarize, the main relations in generalisation are: topology, relative orientation, distance, alignment, density, belonging to a group, being decomposed into (e.g. a city is decomposed into a street network and building blocks, a building block is decomposed into buildings). The main properties are: orientation, isolation, shape, size, granularity.

3 Knowledge Representation Framework

In this section, we present an overview the unified knowledge representation framework for surveillance videos namely CCTV Ontology as depicted in Fig. 2.

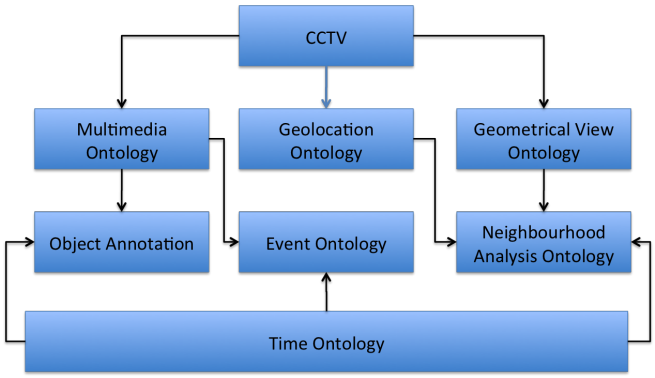

Figure 2. Unified Ontology Framework Model for Surveillance

CCTV Ontology: The CCTV ontology presents an unified platform for integrating multiple ontologies. As opposed

to ontology modeling reported in the literature, which considers either human centric or media centric perspective, we consider modeling semantics from CCTV perspective. In order to represent the completeness of the metadata, this ontology contains the specifications related to CCTV installation. Knowledge related to CCTV includes the manufacturer's details, an uniqueID for identifying the CCTV, a person in charge of managing and maintaining the video footage generated by a particular CCTV. The product details of the CCTV camera are stored with the help of product ontology. The information embedded into the ontology includes, manufacturer's name, date, shipping delivery, installation date and time. Similarly, the details corresponding to the person in-charge would be described using the person ontology [1]. The person ontology provides capability to represent a person's primitive information about the name, details and contact information. In addition to these ontologies, functional properties namely, "hasView", "isIntalledIn', "hasRecorded" and "isStoredIn" imports external ontologies for integration with the CCTV ontology. The CCTV Ontology platform integrating different ontologies is categorized into those ontologies which are already developed, those ontologies which need to be restructured, enhanced and adapted and finally the novel ontologies. An overview of the ontology structure is presented in Fig. 3

Time Ontology: The time ontology adds temporal metadata to the media item that are being recorded for a long time. Usually CCTV camera records 24 x 7 generating a large amount of memory. A typically installation of a CCTV would include either an inbuilt SD storage element or wireless transmitting the data to be stored in a remote server. In either case, the data generated from the CCTV is huge and constantly gets stored in mass storage devices. To facilitate the identification of the media items, the time ontology is used to index media items in terms of CCTV operational time line, as opposed to the video storage time line. To facilitate the ideal representation, functional properties namely "hasBeginning" and "hasInterval" would be used. The time ontology from http://www.w3.org/TR/owl-time/ will be used for importing and integrating in the unified ontology framework

Geospatial Ontology: The Geospatial ontology provides representation of physical location of CCTV in the cartography of a city/region/location. The ontology provides infrastructure to store GPS location along with indexed list of places which could then be mapped onto maps. The information extracted from the combination of geometrical and geospatial ontologies are used to represent other CCTV cameras present in the region through the use of functional property **"hasNeighbourhood"**. The relationship modeled will include different directions or orientations of spatial map namely, east, west, north and south.

Multimedia Ontology: Multimedia ontologies stemming from MPEG-7 representation has been well established and among the semantic implementation of the MPEG-7 descriptions DOLCE based modeling of MPEG-7 has gained the acceptance of the semantic community. The multimedia ontology in general provides infrastructure for representing a media item to be further combined with domain ontology. The multimedia ontology acts as a storage and bridging element to integrate the extracted knowledge with the object annotation ontology and event ontology as discussed below.

Geometrical Behavioural Ontology: This Ontology describes the mapping of 2D video frames to that of physical dimension correspondence. This ontology translates distances from the physical world into pixel values of the videos/images captured. The geometrical spatial ontology, also enables representation of object tracking. The movement of objects at different speeds are being modeled and their physical correspondence is mapped to the video/images. The behavioral ontology enables knowledge representation of objects speed and trajectory.

Object Annotation Ontology: In this paper, we distinguish between object annotation and event ontology on the basis that object annotation provides infrastructure for storing and representing image blobs extracted from the multimedia analysis. The blobs could be of any size and hence it is critical to correspond the image blobs with that of geometrical and geospatial ontology. The object annotation provides a set of pre-defined set of concepts for which analysis algorithms could assign classification categories.

Event Ontology: Event ontology is used to unique represent high-level semantic events such as "car changing lane", "breaking car glass", etc. This ontology presents a methodology for integrating object annotation ontology resources with that of time ontology to create a unique definition of an event ontology. Events ontology is used to populate knowledge extracted from video analysis that correspond to an event timeline. This timeline used in the video is different from the time ontology used to represent the surveillance footage in general.

4 Conclusion and Future Work

In this paper, we have presented an unified knowledge representation framework for modelling, indexing and retrieving information from surveillance videos for forensic applications. The proposed framework facilitated the formalization of contextual, media and event related semantic data as well as providing a platform for the analysis of human language-based queries. Besides the proposed framework is available both explicitly (in the form of media item) and implicitly (in the for of geo-spatial location and neighbourhood awareness). For the creation of such unified platform, several novel ontologies have been proposed, i.e.Geometrical Behavioural ontology, as well as some others have been enlarged and adapted to enable the study of more sophisticated and specific surveillance use cases. Finally, the knowledge representation framework provides a light weight ontology implementation capable to facilitate the query analysis for specific application scenarios.

The continuing future work will focus on creating a repository of resources that is specific for an use-case to analyse the performance of the proposed unified knowledge representation framework in the analysis of a realistic scenario.

[1] http://daml.umbc.edu/ontologies/ittalks/person#

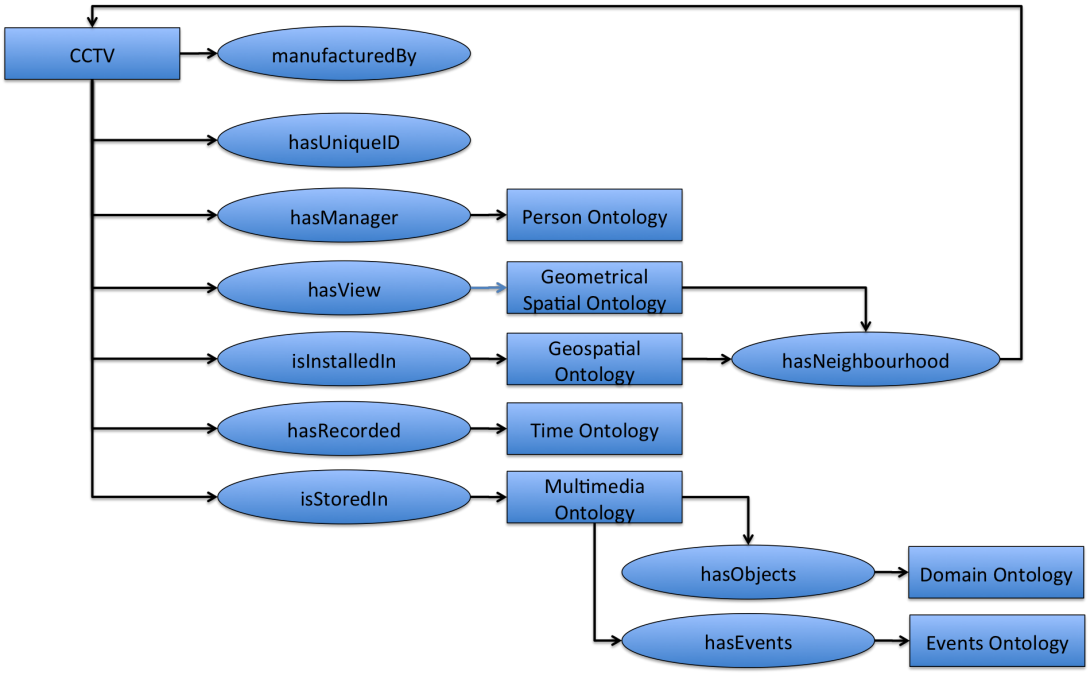

Figure 3. Unified Knowledge Representation Structure for Surveillance Videos

References

[1] N. Anjum and A. Cavallaro. *Intelligent Video Surveillance: Systems and Technologies*, chapter Multi-camera calibration and global trajectory fusion. Number ISBN: 978–1–4398–1328–7. CRC Press, 2009.

[2] N. Anjum and A. Cavallaro. Localization of distributed wireless cameras. In *Proc. of IEEE Int. Conference on Distributed Smart Cameras (ICDSC)*, Como, Italy, 30 August–2 September 2009.

[3] V. Fernandez Arguedas, Q. Zhang, K. Chandramouli, and E. Izquierdo. Multi-feature fusion for surveillance video indexing. In *12th International Workshop on Image Analysis for Multimedia Interactive Services*, 2011.

[4] D. Burghardt, S. Schmid, and J. Stoter. Investigations on cartographic constraint formalisation. In *10th ICA Workshop on Generalisation and Multiple Representation*, 2007.

[5] S. Calderara and R. Cucchiara. Multimedia surveillance: Content-based retrieval with multicamera people tracking. In *ACM International Workshop on Video Surveillance & Sensor Networks*, Santa Barbara, California, 2006.

[6] K. Kaneiwa M. Iwazume and K. Fukuda. An upper ontology for event classifications and relations. In *Lect. Notes Comput. Sci*, 2007.

[7] T. Le and M. Thonnat. Surveillance video indexing and retrieval using object features and semantic events. In *International Journal of Pattern Recognition and Artificial Intelligence, Special issue on Visual Analysis and Understanding for Surveillance Applications*, 2009.

[8] Y. Liu, R. McGrath, S. Wang, and M. Pietrowics J. Futrelle J. Myers. Towards a spatiotemporal event-oriented ontology. In *Microsoft eScience Workshop*, December 2008.

[9] Y. Ma and I. Cohen. Video sequence querying using clustering of objects appearance models. In *International Symposium on Visual Computing*, 2007.

[10] W. A. Mackaness. Understanding geographic space, in generalisation of geographic information: Cartographic modelling and applications. In *Mackaness and Ruas (eds), Elsevier.*, 2007.

[11] A. Sheth and M. Perry. Traveling the semantic web through space, time and theme. In *IEEE Internet Comput*, December 2008.

[12] I. Tziakos, A. Cavallaro, and L.-Q. Xu. Event monitoring via local motion abnormality detection in non–linear subspace. *Neurocomputing*, page to appear, 2009.

[13] S. Velipasalar and L. M. Brown. Detection of user-defined, semantically high-level, composite events, and retrieval of event queries. In *Multimedia Tools and Applications*, 2010.

[14] C. Zhang and X. Chen. Semantic retrieval of events from indoor surveillance video databases. In *Pattern Recognition Letters*, 2009.