ISIS Data Model: Towards a Common Event Model

Behrang Qasemizadeh, Ian O'Neill, Philip Hanna, Darryl Stewart

Institute of Electronics, Communications and Information Technology (ECIT)
Queen's University of Belfast
Belfast, BT3 9DT, Northern Ireland
qasemizadeh@gmail.com, {i.oneill, p.hanna, d.stewart}@qub.ac.uk

Abstract. This paper describes a data model for content representation of temporal media in an IP based sensor network. The model is formed by introducing the idea of semantic-role from linguistics into the underlying concepts of formal event representation with the aim of developing a common event model. The architecture of a prototype system for a multi camera surveillance system, based on the proposed model is described. The important aspects of the proposed model are its expressiveness, its ability to model content of temporal media, and its suitability for use with a natural language interface. It also provides a platform for temporal information fusion, as well as organizing sensor annotations by help of ontologies.

Keywords: Data Model, Content Modelling, Ontology.

1 Introduction

Because of advances in computer network technology, exploration of the internet, and the decreasing cost of hardware, sensors have been increasingly adopted in several information infrastructures for different applications such as information collecting, modelling, and retrieving. [1] An acute research community is currently focusing on surveillance sensor networks in applications like security, and fighting crime. Key research challenges in this domain are fusion of heterogeneous information in sensor networks (and even homogenous information such as multi camera surveillance), in addition to providing easier access to stored information.

This paper suggests a model for content modelling of temporal media in a surveillance sensor network. The proposed data model aims to develop a platform for information fusion, and pushes the boundaries one step towards the goal of a *common event model*. Westermann and Jain have discussed the notion of *common event model* in [2]. With emphasis on the importance of a *common event model,* they enumerated important features that can be achieved as an outcome of having a common event model, among them common base representation, unified media indexing, as well as common event retrieval and mining environments.

The proposed model tries to realize the mentioned goals by use of ontologies and applying the idea of semantic-role from linguistics to the underlying concepts of formal event representation. In the proposed model, the outputs of sensors in time instants are tagged by ontology-provided vocabulary, and events in temporal media are mined by means of temporal logic and the time-invariant relation between sensors of the network. As a simple example, if the distance relation between two stationary fingerprint readers is known, then reading the same fingerprint at these two different fingerprint readers, at two different time points, means that a person moved from the location of the first fingerprint reader to the location of the second fingerprint reader.

The remaining sections of this paper are organized as follows: section 2 proposes a data model for content representation of temporal media like surveillance video data; a prototype system that uses the data model and the data flow within the system is explained in section 3; finally, we offer conclusions in section 4.

2 Proposed Data Model

Figure 1 shows the proposed model. The model has three elementary data types namely Property, Object (entity) and Event. Data elements hold values that correspond to the vocabulary introduced by the ontology/ies for that data element. Furthermore, each data element may relate to another data element through a semantic/thematic role. A Time Ontology supports the temporal aspect of the model such as temporal granularity, i.e. how often the model is refreshed by inputs of sensory devices, as well as temporal metrics.

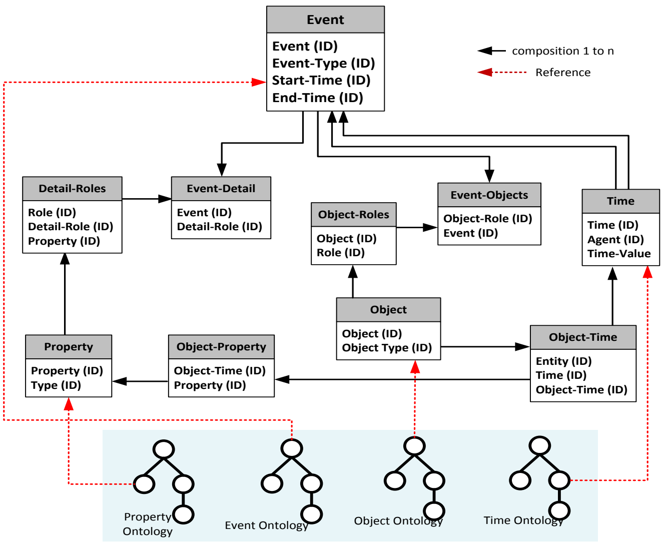

Figure 1. The Proposed Data Model.

The two most important views of the data scheme are *Event* and *Object*. *Event* is a constituent for representing actions e.g. approaching, coming near or nearer, in space or time. *Object* refers to “things” or entities that we identify in a domain of interest, e.g. in an office surveillance model, objects may include persons, and stationary items such as computers. *Property* refers to the qualities of objects, or is used to describe an event through a semantic role. For example, location can be a quality assigned to *Objects* for specific time, or it can be a factual datum that completes the meaning of an action like "approaching a location". In a domain of interest, there might be more than one *Property*; in this case, each *Property* will be described by an individual ontology of that *Property*.

In the proposed model, each instant output of a sensor is uniquely tagged by the vocabulary provided by *Object* and *Property* ontologies, and is accompanied by a temporal tag. The temporal tag uniquely identifies the source of information i.e. a sensor device, and its modality; moreover, each temporal tag has a pointer to real sampled data from a sensor. For example, a temporal tag for a surveillance camera identifies one camera in a multiple camera network. Moreover, the temporal tag provides a pointer to the video frame that has been captured, at that time instant, and by that camera; e.g. the pointer can be a URL of a jpeg image file.

Once the model provides a common vocabulary for annotating the output of sensors, i.e. a vocabulary provided by ontologies, it is possible to check the output of sensors against each other by help of defined relations within ontologies. The checking procedure can be employed for assigning a confidence measure, and/or for discovery of anomalies. Because the output of sensors is annotated by ontology-provided vocabulary, the check rules for data consistency can be written for concepts introduced by ontologies, rather than for each individual sensor.

As mentioned earlier, another distinct feature of the proposed model is the use of semantic-role [3] in its structure. As *Event-Objects* and *Event-Details* in figure 1 show, *Object* and *Property* are related to *Event* through a composition of semantic-role labeled entities. The introduction of semantic-role into the model plays two roles. First, it maintains relationship between concepts which are defined in two different ontologies e.g. between concepts in an *Object Ontology*, and an *Event Ontology*, in other words intra-ontology relationships between concepts; Second, semantic-role labels provide linguistics knowledge about how to interpret and map factual data to/from natural language utterances.

To explain the importance of the semantic role, we continue with an example. Video Event Representation Language (VERL) [4] is a formal language for video content Modelling. VERL is based on first order logic and describes an ontology of events; individuals may define their own event ontology in a domain of interest and exploit VERL to describe that event ontology. In the VERL framework, each video instance is accompanied by a mark-up in Video Event Markup Language (VEML); VEML describes the content of a video stream according to its companion VERL. In this matter, our work has benefited from the underlying logic behind the VERL framework and the relevant event detection procedure; however, the proposed approach take advantages of ontologies describing the background knowledge of the domain, and it uses the definitions of events and their semantics in the event ontology to go one step further by introducing semantic-roles into the model proposed by a formal language like VERL.

The VEML annotation for a sample *approach* event is shown below. The *approach* event has a certain meaning encoded in rules and conveyed in the VERL ontology. The definition of the *approach* event holds two arguments, in addition to other details such as the start frame and end frame for a certain instance of approach event in a specific video stream, as follows, (example taken from [5]):

```
<event type="APPROACH" id="EVENT1">
  <begin unit="ms">136</begin>
  <end unit="ms">147</end>
  <property name="name" value="P1-approaches-door"/>
  <argument argNum="1" value="P1"/>
  <argument argNum="2" value="DOOR"/>
</event>
```

The VEML representation of the sample *approach* event above implies the statement "P1 approached the Door" in a human observer's mind and it is encoded in the definition of *approach* in VERL; however, to enable machines to have such an interpretation from the above annotation, we need a formal description which tells a machine how to interpret/translate that annotation to/from natural language. This can be done with the help of semantic-roles.

If we introduce the first argument of an *approach* event as the *agent* of the event and the second argument as the *goal* of the event, then we are certainly able to map an utterance like the above statement into/from its companion VEML representation. The following shows our suggested XML representation for the first and second arguments of VEML representation:

```
<argument semantic_role="agent" value="P1"/>
<argument semantic_role="goal" value="DOOR"/>
```

Because VEML is a formal language, it is possible to write unambiguous ontological mappings from the VEML representation into the proposed model where we know the semantic role of the arguments. In effect, the above XML representation will be encoded through a set of facts organized around the elements of data model. To give more insight, the next section describes the architecture of a prototype system.

3 Prototype System

The proposed data model has been employed in a prototype system of a doorway surveillance system (Figure 2). The system automatically captures video from multiple sources and annotates the video with the gender of people who walk the doorway and enter a controlled environment. The system comprises three main components: a sensor-based analysis component, in this case cameras in addition to their companion Image Analyzers (IA), a Data Manager (DM), and an Event Detection (ED) component. System components are implemented as autonomous agents communicate through TCP/IP connections.

IAs identify people and their location, as well as their gender, and assign them a unique ID. Detailed steps of data flow are as follows (Figure 3): (1) an IA lets the model know about the temporal granularity, i.e. time instants that are used to refresh the information here a single camera creates a new temporal tag containing an identifier for the sensor and a pointer to the captured data at that time; (2) it creates an instance of a new object (if it does not already exist); (3) it lets the model know at what time instant this object was seen (or sensed); (4) it writes the detected values for the object's properties – like location and the gender of person for a camera– in the form of *Object-Property* facts. Needless to say, the Data Manager corroborates data flow each time a component submits new facts, and assigns them a confidence value.

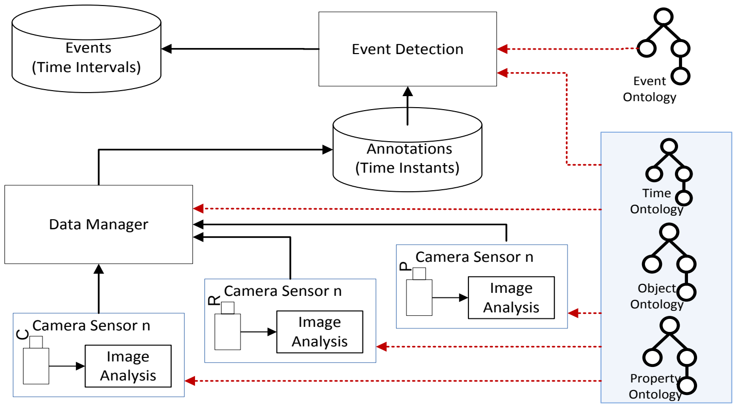

Figure 2. Block diagram of the prototype system. Sensors (here IP-based surveillance cameras with their companion image analyzers) annotate observations with the help of vocabulary provided by *Time*, *Property*, and *Object* ontologies, and leave annotations in sub part of the data model. The Data Manager checks data aggregation and assigns confidence measures to annotations. Event Detection mines events from annotated observations.

Simultaneously, whenever sensors refresh the model, they send an XML message to the ED. The ED is implemented in Prolog. The message to the Prolog component contains information about temporal granularity, time order, and profile of the sensor device that refreshes the model. The ED uses temporal logic to infer atomic events from the provided annotations. Annotated time instants and inference rules are designed to assign the semantic-role of the objects and the properties of events. The right-hand section of Figure 3 shows the data flow for asserting an inferred event into the data model.

4 Conclusion

This paper introduces a data model for content Modelling of temporal media in a sensor network, like a multiple camera surveillance system. To further the goal of establishing a *common event model*, an ontology supported data model connects data elements by means of semantic-role labeled relations. Our aim is to show that a complex linguistic structure assists the representation of deep semantics. Rules based on the vocabulary introduced by ontologies help check data aggregation and consistency, independent of physical sensor devices. Introducing semantic roles in an event Modelling framework provides a mean of systematic mapping of outcome of semantically labeled natural language constituents into elements of the data model and vice versa. Moreover, semantic-role relations can be used for managing intra-ontology semantic relations, i.e. semantic relations between concepts that are defined in different ontologies.

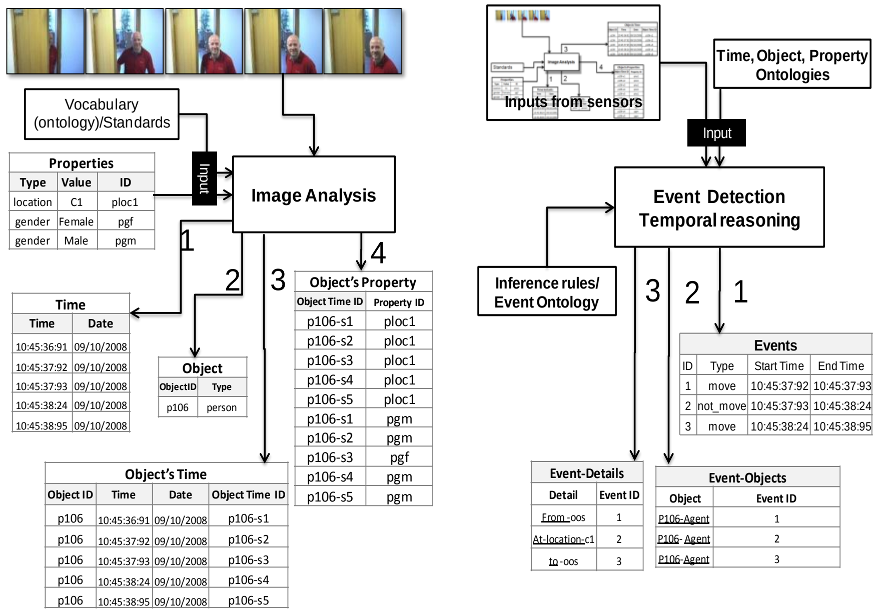

Figure 3. Schematic data flow. The left-hand side shows the order of asserting a sensor's observations (e.g. an IP camera with its companion image analyzer) in the data model; event detection is shown on the right-hand side.

Acknowledgment

This work is supported by the EPRSC under project number EP/E028640/1 (ISIS).

References

1. Sheth, A. Henson, C. Sahoo, S. S.:Semantic Sensor Web. IEEE Internet Computing, (2008).
2. Westermann, U., Jain, R.: Toward a Common Event Model for Multimedia Applications. IEEE MultiMedia (2007).
3. Jackendoff, R. S.: Semantic Structures, MIT Press, (1992).
4. Francois, A. R., Nevatia, R., Hobbs, J., Bolles, R. C.: VERL: An Ontology Framework for Representing and Annotating Video Events. IEEE MultiMedia. (2005)
5. Bolles, B., and Nevatia, R.: ARDA Event Taxonomy Challenge Project, Final Report, (2004).