Towards a Common Event Model for an Integrated Sensor Information System.

Chris Fowler[1] and Behrang Qasemizadeh[2]

[1] ECIT Institute, Queen's University Belfast, Northern Ireland Science Park, Queen's Road, Belfast, UK.
[2] DERI, Unit of Natural Language Processing, National University of Ireland, Galway IDA Business Park, Lower Dangan, Galway, Ireland.

Email: c.fowler@qub.ac.uk, behrang.qasemizadeh@deri.org

Abstract. This paper describes steps towards a common event model for event representation in sensor information systems. The model builds on a representation of events, and introduces the idea of *semantic-role* from linguistics, attaching semantic annotations to the underlying concepts of formal event representation. We describe how semantic-role annotations facilitate linkages across the descriptive layers drawn from sensor data ontologies, allowing us to reason more effectively about *causality* within a sensed environment. An overview of the parent system for which the event model was derived is given to provide application context for the work. This is followed by a detailed description of the model, together with a description of how the model is used within a prototype system that illustrates the model in practice. The paper ends with conclusions and issues for further work.

1 Introduction

The vision of a *semantic reality*, as described by Manfred Hauswirth in 2007 [5], posits a world where sensor technology and the semantic web combine to enable a single unified view that bridges the gap between virtual and physical space. The result would be a machine readable semantic layer, rooted in an ontological domain-description, making possible a machine navigable information-web that mirrors reality. The use of ontologies provides a vocabulary and classification mechanism through which specific domains may be described. These descriptions are encoded in meta-data used to annotate the data gathered from the sensor-web. The addition of a semantic layer enables the possibility to reason about and understand the relationships between the '*things*' described in the ontology-based annotations. The possible benefits and applications for this machine-readable real-time virtual lens on the world are numerous: health-care provision, security and crime prevention, traffic management, wild-life preservation, environmental monitoring, are only a few such examples. Of course, the potential harmful uses are equally present, but we should not let this prevent us from exploring the issues and challenges towards this new technology.

The true semantic reality is some way off. If we consider, however, that the semantic reality envisioned could, if viewed from a different perspective, be described as a collection of intersecting *semantic sensor webs* [10], then we are closer than we think.

Taking Sheth and Henson [10] as a datum, a semantic sensor-web is seen as a network of remote, autonomous sensors, which detect changes (events) in the environment and make these data available as an information source on the Web. These source data may then be used for information-fusion towards a higher-level understanding of the sensed environment; for example, weather tracking, flood monitoring, avalanche prediction, or crime prevention. Each sensor-web may include a number of different modalities. The addition of a semantic layer allows for a richer interpretation of the source data.

The Integrated Sensor Information System project (ISIS [1]) maybe viewed as an example of a semantic sensor-web. Its application is crime prevention on public transport. Its fundamental structure is a distributed sensor-web, with both remote and central semantic-based analytics capability, designed to fuse sensor data towards an understanding of the environment under surveillance from a security and crime prevention perspective - so called *situation awareness*. ISIS is designed to: assert threat levels on public transport using embedded sensor-array nodes positioned on-board buses as they traverse the transport network; inform key decision makers of changes in threat-level via a control room interface; and manage its own network. It is an example of applying semantic senor technology in a real-world domain.

A key aspect of ISIS is the use of multi-modal sensors (video cameras, microphones and radio-frequency sensors). Because we are using different modes of sensor, each type will '*speak*' a different language. To make sense of this, we must unify the differences towards a *common language*, in order that the system as a whole may be mutually understood. This requirement is not unique to ISIS and the common approach is to use ontologies to markup identifiable objects and events as they occur in the data. (We have defined a set of ontologies for this purpose). In addition to marking-up objects and events, we also want to be able to reason about the causality of events and their relationship to the objects involved. We have achieved this by introducing the notion of *semantic-role* from linguistics.

The introduction of semantic-role allows us to define intra-ontology-relations as a common platform for event modelling and causality inferencing. The semantic-role-relations provide us with the logical linkages we need between the different elements of the data-model (see figure 1). They allow for a formal interpretation of the different relationships between the informative elements defined in different classes of ontology. Consider this example, "a *man approaches* the *chair*": in this case we would assign the semantic-role *agent* and *goal* to the man and chair respectively. (The vocabulary for describing the *goal* and *agent* are taken from the domain ontologies, as are the *event* descriptions). This structure (*agent-event-goal*) may, at a given time, and within the rules of the ontology governing the event, be evaluated using inference rules, and said to be a true/false state-

ment. The concepts within each class of ontology describing a specific object-, or event-class, may now be related by assigning the semantic-role relation to the participant objects of an event. This allows us to reason about events and search for specific events involving specific objects identities. The semantic-role relations between different object and event ontologies in our model, therefore, link the objects in a scene to the events in a scene across different sensor-modalities. In this way they act as a semantic-bridge, linking the knowledge-base of the domain. Reasoning-agents are then able to navigate the information-space via formally defined semantic relationships. The links between the descriptive layers that focus on the objects and properties, and the layers describing the events within a scene, can be now be used to determine *who did what*. This is not sufficient for our purpose; we also need to know the *when*.

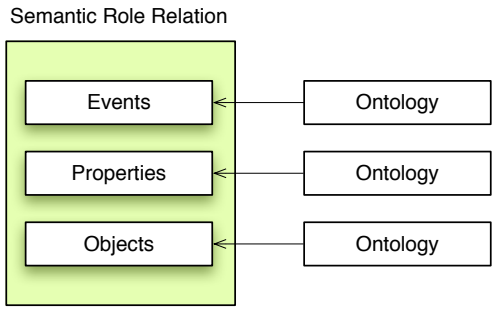


Fig. 1: Semantic Role Relations.

Events do not happen without *time*. We must therefore include time in our modelling process. Events happen over time intervals, therefore, synchronised data streams are needed to determine unique object-event-relations as viewed from different sensors.

In ISIS, sensors, together with their associated analytics, identify objects in the sensed data from multiple sources. These data need to be unified at a particular time instance across these multiple sources. For example, a noise heard by an audio sensor may be matched in time with a video data of the same scene that shows a person dropping a plate; we may then say that these two sensors sensed the same event in time; and say: who dropped the plate! Another key addition to time is the notion of *space*. An array of sensors, fixed to survey a bounded environment, are, by their physical location, co-located in space. If their sensor-view is aligned in both time and space, then we are able to infer that they are 'watching' the same scene.

Additional contextual information can also be used to unify a scene. Consider this example: if a distance relationship is known between two finger-print readers, and at two different time-instances the same finger-print scan is read at each, we can infer that in the time-interval between readings the person (or

at least their finger) moved from the location of one reader to the location of the other. This example begins to show the rich set of information that comes together to create understanding, and shows the links between objects, events, sensor activity, physical space, time, and causality.

To begin the process of building a machine readable semantic-skin over the sensor data captured by the ISIS system, we must define the fundamental elements of our data-model. It must be capable of capturing the objects, properties, events, time and space, as well as the semantic-role-relationship between the different ontology conceptualizations. To do this we adopt the same principals outlined by Westermann and Jain [11] for a *Common Event Model*.

This paper describes our interpretation of Westermann's model towards a common event model for ISIS. Our model may be seen as the rich descriptive *skin* that wraps different layers of abstraction within multiple sensor data sources. At its core are the unique objects in the scene, captured at each time instant, with additional layers that describe atomic events, low-level events, higher-level events and ultimately domain-specific behaviors, each occurring over increasing time intervals. By linking ontologies across each layer using the notion of a semantic-role-relation, our model allows for greater understanding of causality within the sensor data towards an integrated sensor information system. Our work uses the Video Event Representation Language (VERL) and the Video Event Markup Language (VEML) presented by Fraccois et al [4] and the theory of *Causality* from Hobbs [6] as its base.

This paper is structured as follows: section 2 presents an overview of the ISIS system that places our work in context. Section 3 gives a full description of the Common Event Model for ISIS, together with database representation, example event annotations. A prototype system developed to test and explore issues with the work is described in Section 4, with details of how our model integrates the elements of the system. Conclusions and further work are presented in section 5.

2 ISIS System Overview

In this section we present a high-level view of the ISIS system. Although the application for our system is crime prevention, we believe that by making a separation between the sensor hardware infrastructure and the language layers through which the system represents and interprets the sensed environment, ISIS can be applied to many different domains.

Figure 2 shows a high-level view of the ISIS system. Four key elements exist:

1. *A remote sensor-array node*. In our application this is located on-board a bus traversing the transport network. Its main function is to sense the scene and detect in the data the profile and mix of passengers on-board, and infer any domain specific behaviors relating to security and crime detection. This real-time risk inferencing contributes to an on-board risk level. Once the risk

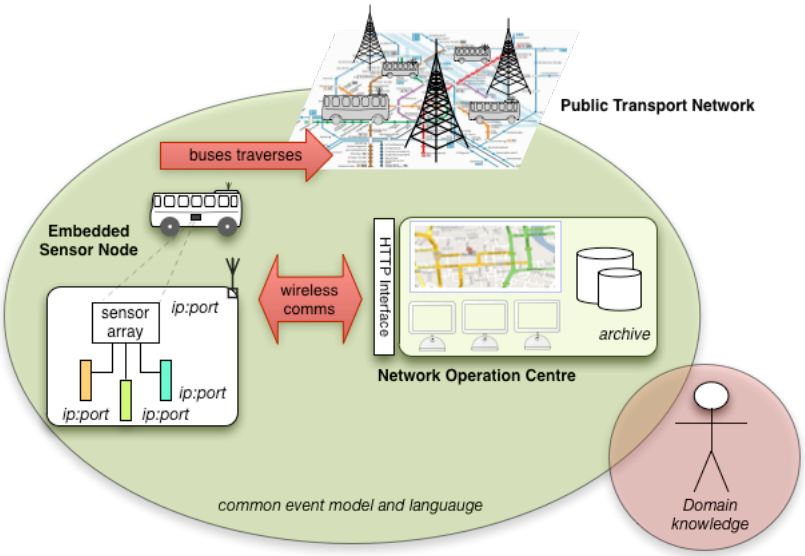


Fig. 2: Separation of Concerns in the ISIS Infrastructure

level rises above a certain threshold, an alert message is communicated to a human-operator in a network control centre.

2. *A wireless communications infrastructure.* Vehicles traversing the transport system communicate via message exchange with a control room. This is done over a wireless network.
3. *Network Control Centre.* The control centre is manned by human operators - domain experts. It has two main functions: a) to provide real-time visualisation of the current state of the sensor-array network, allowing operators to respond to alert messages as they are triggered, and b) archive and retrieval capabilities for storing data and gathering evidence in the event of a crime. The human operator is an intimate part of the system. Their domain knowledge is a bridge between the events in the context scene - detected, annotated and stored in the archive - and the world-view of witnesses, or other interested parties. As such, the vocabularies used should reflect the domain under scrutiny (in our case, crime and security). By using ontologies we are able to semantically map queries from an operator to queries over the data archive.
4. *Common Event Model and Language.* Key to unifying the ISIS infrastructure is a common event model, capable of capturing the physical environment being surveyed in terms of time and space as well as the objects and relationships within the scene. The ontological language used to describe each scene must be shared across all participants in the system - human and process. Providing this common language platform has proved a key enabler

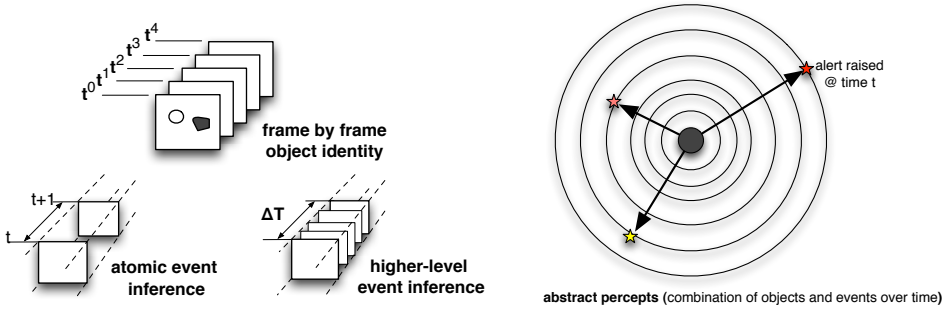
for human interaction with the system when retrieving specific events from the sensor data archive. It also provides the necessary separation of data-model from hardware infrastructure. As new domains are added to the event model ontologies and mapped on to rules within the system model, so new behaviors and events may be monitored.

We now go on to discuss the breakdown of sensor data towards the fundamental constructs of the common event model proposed.

2.1 Perceiving the Physical Environment

In the words of Marvin Gaye "*The World is just a great big Onion*" [2]. This is a view we take when perceiving the world via the ISIS sensor-web towards the goal of triggering an alert of a specific security/crime event on-board a bus in our network.

Figure 3 shows two views on the data. Figure 3a shows how the data stream is broken down into individual frames (in reality these may be key-frames rather than every frame). At a time instant we detect within the frame the identifiable objects and their properties. The objects are assigned a unique identifier. To determine events, we must examine the difference between frames over a time interval. At the lowest level - atomic-event - this is done with consecutive frames. For higher-level events a great time interval is used, as are more frames. Taking this approach to determining events, we can see that over time, layers within the data appear that correspond to the activity within the scene. This is the approach we use to trigger an alert within the system. By considering domain specific behaviors as a collection of related events, we are able to determine at what point a set of events may be perceived as a '*looked-for*' behavior. At this point the system will raise an alert. This is illustrated in figure 3b.



(a) Frame analysis of Sensor Data (b) High-level Event Triggers

Fig. 3: Abstract layers and frame instances within sensor data.

To achieve this requirement we need a formal language to describe the sampling of the real-world discussed above, and to represent this in a conceptual model. This language and process of representation is now discussed.

3 Towards a Common Event Model

Figure 4 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 the temporal granularity, i.e. how often the model is refreshed by inputs from sensory devices, as well as temporal metrics.

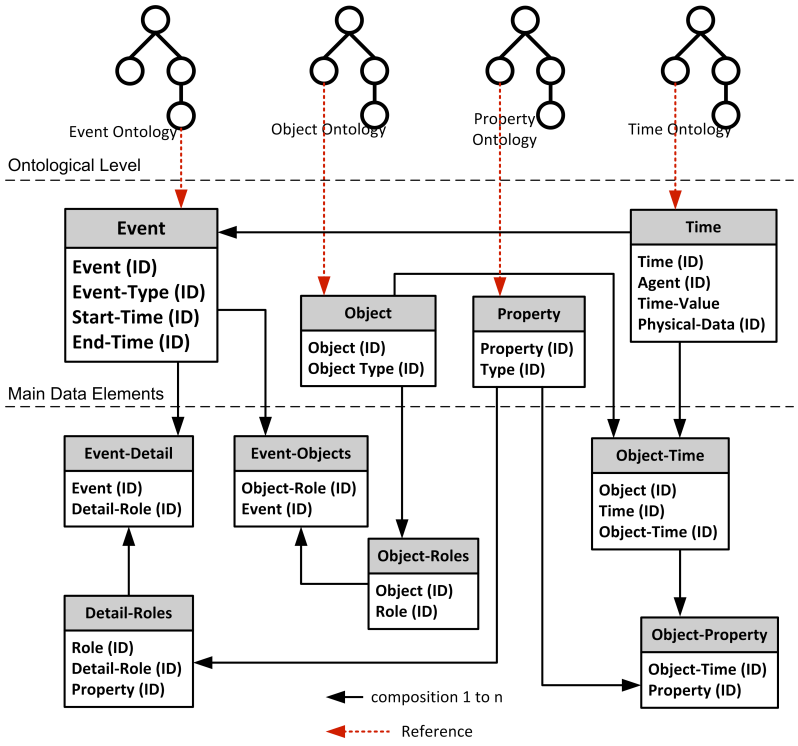


Fig. 4: Common Event Model for the Integrated Sensor Information Systems

The two most important views of the data scheme are Event and Object. The Event is a constituent for representing actions e.g. approaching, coming

near or nearer, in space or time. The Object refers to *things,* or entities, that we identify in a domain of interest, for example, in an office surveillance model, objects may include persons, and stationary items such as computers or desks. The 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 a 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 the Object and Property ontology, and 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 data sampled by a sensor. As an 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 - a pointer can be a URL of a jpeg image file.

As the model provides a common vocabulary for annotating the output of sensors, it is possible to check the output of sensors against each other by defined relations within the ontologies. A checking procedure can then be employed - whether for assigning a confidence measure, and/or the discovery of anomalies - allowing the checking-rules for data consistency to be written for concepts introduced by ontologies, rather than for each individual sensor. This ability separates the language of description and inference from the sensor hardware infrastructure.

As mentioned earlier, another distinct feature of the proposed model is the use of semantic-role [7] in its structure. As figure 4 (Event Objects and Event Details) shows, 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 major roles: firstly, it holds a relation between concepts which are defined in two different ontologies, e.g. between concepts in Object Ontology, and Event Ontology, forming an intra-ontology relationship between the distinct concepts, and 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 semantic role, we continue with an example. The Video Event Representation Language (VERL) [4] is a formal language for video content modeling. VERL is formed on first order logic to describe an ontology of events; individuals may then define their own event ontology in a domain of interest and exploit VERL to describe that event in an ontology. In the VERL framework, each video instance is accompanied by a Video Event Markup Language (VEML) tag [3] - 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 relevant event detection procedures. In addition our proposed approach takes advantage of ontologies

in the supported domain's background knowledge, 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.

A VEML annotation for the sample "approach" event is shown below (example 1). The approach event has a certain meaning encoded in *rules*, conveyed by the VERL ontology. The definition of the approach event holds two arguments (argNum1 and argNum2) each with a corresponding value. In addition, other details such as the start frame and end frame for a specific instance of approach event in a specific video stream, as well as a name for the event. This complete VEML annotation refers to a specific event instance.

Example 1. VEML Approach Event

```
<event type="APPROACH" id="EVENT1">
    <begin unit="ms">136</begin>
    <end unit="ms">147</end>
    <property name="name" value="Person1-approaches-DOOR1"/>
    <argument argNum="1" value="Person1"/>
    <argument argNum="2" value="DOOR1"/>
</event>
```

The VEML representation of the approach event above implies the statement "*Person1 approaches Door*1" in a human observer's mind and is encoded in the definition of "approach" event in the VERL rule ontology. To enable machines to have such an interpretation from the above video annotation however, we need a formal description, which tells a machine how to interpret/translate the VEML annotation to/from natural language. (We say natural language here as this refers to the expressiveness of the proposed model - this is emphasised by Westermann and Jain [11]) - this expressiveness requirement can be achieved by the help of semantic-role.

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 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 (example 2):

Example 2. XML Representation introducing Semantic Role

```
<event type="APPROACH" id="EVENT1" begin=''T03'' end=''T07''>
    <argument semantic_role="agent" value="Person1"/>
    <argument semantic_role="goal" value="DOOR1"/>
</event>
```

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 each argument. In effect, the above XML representation will be encoded through a set of facts organized around the elements of the data model. To give more insight, the next section describes the architecture

of a prototype system that uses the event model described above, to integrate the elements of a doorway surveillance system.

4 Prototype System

The proposed data model has been employed in a prototype system for a doorway surveillance system (see figure 5). The system automatically captures video from multiple sources and annotates the video, identifying people as well as their gender property as they walk and enter into a controlled environment.

The system comprises three main components: a sensor based analysis component (shown as camera sensors and their companion Image Analyzers (IA)), a Data Manager (DM), and an Event Detection (ED) component. The system components are implemented as autonomous agents communicating through TCP/IP connections.

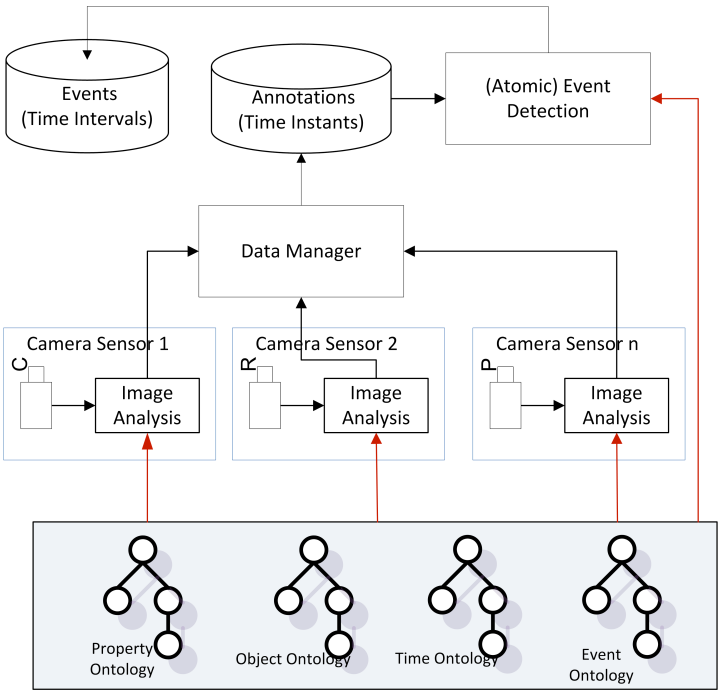


Fig. 5: Block Diagram of the Sensor-based Prototype System.

The Camera Sensors are annotating observations using vocabularies provided by the time, property, and object ontologies and writing the annotations to a sub-part of the data model. The Data Manager checks data aggregation and assigns confidence measures to annotations. The Event Detection process mines for

events in the annotated observations and writes these to another sub-part of the model. The Image Analysers identify people and their location, as well as their gender, and assign them a unique ID. This is done by mapping extracted features using Principal Component Analysis [8,9] to high level concepts described in the ontologies, for example the type of object.

Figure 6 shows how the proposed model integrates the physical aspects of the ISIS sensor network. Referring back to figure 3b it is possible to see how the model integration illustrated in figure 6 produces layers within the data, where each layer is rooted in the information-base described by the pool of ontologies that make up the domain. At the core (1st layer) the atomic events are captured as time invariants. These represent the lowest level of detail infered by the system. Each subsequent layer represents a skin of new, infered knowledge, whose pool of knowledge is drawn from that held by the the previous $n-1$ layers.

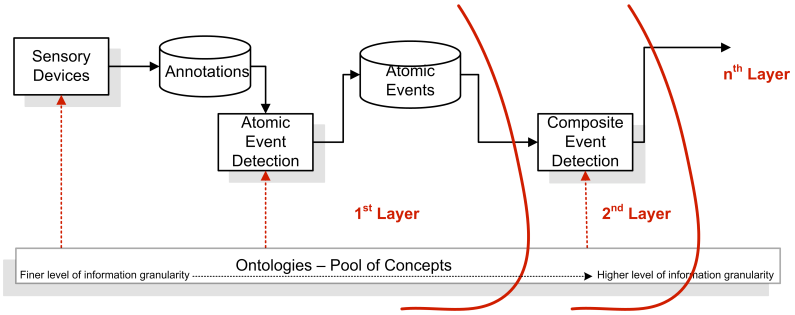


Fig. 6: A layered view on system inferences. At the core of the system, sensors are annotating time invariants.

The Event Detection procedure, as it is described above, may be repeated for several turns. Figure 6 shows this layered view of system inferences. The Atomic Event Detection procedure detects the most granular events. These are then used by the system as higher level abstract definitions for inferencing events at the next level of granularity; this may also be viewed from a temporal granularity perspective. Such a setup for event detection may be helpful when employing different communication technologies for data exchange at the physical network layer, as each communication may refer to specific abstractions captured within the data.

5 Conclusions and Further Work

This paper introduces a scheme for content modeling of temporal media in an integrated sensor network. The aim of the work is to move a step closer towards a

common event model for integrated media data as described by [11]. To do so, an ontology-supported data model that connects data elements using semantic-role-relations was introduced. Our aim was to show that by introducing the notion of semantic-role from linguistics, we are able to better represent semantic content of sensor-data captured within our sensor-web. The use of ontologies aids the checking of data aggregation and consistency towards a unified view of the world under surveillance, independent of the physical sensor devices. Introducing semantic roles in an event modeling framework provides a means for systematic mapping of the outcome of semantically labeled natural language constituents, into elements of a 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. We showed how this model may be used to integrate the elements of an integrated sensor information system, representing infered domain knowledge as layered skins with increasing information granularity.

The current system is implemented in Prolog with ontologies implemented in first order logic. Converting the ontologies to a standard ontology language such as Ontology Web Language is considered for immediate future. Although temporal reasoning and representing temporal inference rules remains untouched, this also forms a part of our future work. In addition, the approach proposed raises an issue regarding the trade-off between the real-time inferencing of events and the storage of events as higher level abstractions used in higher level reasoning. For further experimental study and investigation therefore, is the balance between the granularity of the stored events and those infered in real-time.

6 Acknowledgement

The authors would like to thank Dr Jiali Shen of ECIT, Queens University Belfast, for his work towards object identification and gender profiling in video data. The work presented in this paper is supported by the Integrated Sensor Information System (ISIS) Project, funded by the Engineering and Physical Science Research Council, reference number EP/E028640/1.

References

1. ISIS - An Integrated Sensor Information System for Crime Prevention.
2. N. Ashoford, V. Simpson (performed Marvin Gaye, and Tammi Terrel). The onion song. In *Easy*, number TS294. Tamla, 1969.
3. Bolles B. and Nevatia R. ARDA Event Taxonomy Challenge Project, Final Report, 2004.
4. Alexandre R.J. Francois, Ram Nevatia, Jerry Hobbs, and Robert C. Bolles. Verl: An ontology framework for representing and annotating video events. *IEEE MultiMedia*, 12(4):76–86, 2005.
5. Manfred Hauswirth. Semantic Reality - Connecting the Real and the Virtual World. In *SemGrail Workshop 2007*. Microsoft, Seattle, 2007.

6. Jerry R. Hobbs. Causality. In *Fifth Symposium on Logical Formalizations of Commonsense Reasoning*, Common Sense. New York University, New York, New York, 2001.
7. R. Jackendoff. *Semantic Structures*. MIT Press, Cambridge, MA, 1990.
8. Ratika Pradhan, Zangpo Gyalsten Bhutia, M. Nasipuri, and Mohan P. Pradhan. Gradient and Principal Component Analysis Based Texture Recognition System: A Comparative Study. In *ITNG '08: Proceedings of the Fifth International Conference on Information Technology: New Generations*, pages 1222–1223, Washington, DC, USA, 2008. IEEE Computer Society.
9. Sameena Shah, S H Srinivasan, and Subhajit Sanyal. Fast object detection using local feature-based SVMs. In *MDM '07: Proceedings of the 8th international workshop on Multimedia data mining*, pages 1–5, New York, NY, USA, 2007. ACM.
10. A. Sheth, C. Henson, and S. S. Sahoo. Semantic sensor web. *Internet Computing, IEEE*, 12(4):78–83, 2008.
11. Utz Westermann and Ramesh Jain. Toward a common event model for multimedia applications. *IEEE MultiMedia*, 14(1):19–29, 2007.