SEMANTICS-PROVIDED ENVIRONMENT VIEWS FOR NORMALITY ANALYSIS-BASED INTELLIGENT SURVEILLANCE

Lorenzo M. López-López, Javier Albusac, José Jesús Castro-Schez and Luis Jiménez-Linares
Escuela Superior de Informática, Universidad de Castilla-La Mancha, Paseo de la Universidad 4, Ciudad Real, Spain

Keywords: Surveillance system, Intelligent surveillance, Multisensor surveillance, Multi-sensor data fusion, Sensor data integration, Sensor processing and control, Multiagent system.

Abstract: Nowadays, the design and development of intelligent surveillance systems is a hot research topic thanks to the recent advances in related fields such as computer perception, artificial intelligence, and distributed devices infrastructures. These systems are gradually going from the classic CCTV passive surveillance systems towards systems which are capable of offering automatic interpretation of the events occurred in a monitored environment and decision support information based on the data obtained from a number of heterogeneous perception devices. In this work, we introduce the formal definition of an intermediate layer in the architecture of an intelligent surveillance system, of which purpose is to provide the components responsible for performing the reasoning processes with the data from the environment they need. Such data is provided by means of environment views, which are data objects that contain not only data from different sensors, but also associated semantics which depends on the particular context in which the analysis of the normality of a concept is performed.

1 INTRODUCTION

The problem of intelligent surveillance deals with the perception, interpretation, and identification of the activities and situations that occur in a monitored environment.

The technological evolution of surveillance systems started with the *CCTV (Closed-Circuit Television) based surveillance systems* (M. Valera and S. A. Velastin, 2005). These employed a set of video cameras distributed throughout the environment and usually connected to a central security department, where the security personnel must be continuously monitoring what is shown in the monitors. This approach is often referred to as *passive surveillance*.

The second generation of surveillance systems has been possible due to the recent advances in the computer vision and artificial intelligence fields, such as new image processing techniques (L. Rodriguez-Benitez et al., 2008) or behaviour and activity recognition. *Intelligent visual surveillance systems*, as they are commonly known (W. Hu et al., 2004), are able to automatically learn and recognise activity patterns or situations that occur in an environment from the data provided by a collection of video cameras distributed throughout the environment (P. Remagnino et al., 2004) (R. T. Collins et al., 2000). Although these systems suppose a clear improvement over the first generation ones, some of the employed algorithms and techniques are still not mature enough, which results in systems that often need a long time to response and trigger too many false positive alarms (M. Valera and S. A. Velastin, 2005).

Finally, the third generation of surveillance systems refers to those systems which are able to use the information provided by a heterogeneous set of sensors distributed across the environment. Such set of sensors is composed of devices of relatively low cost, such as radio-frequency identification (RFID) sensors, audio sensors, presence detection sensors, and video cameras. As a consequence of this heterogeneity, these systems are able to obtain a more accurate knowledge of the environment (R. C. Luo et al., 2002). However, some aspects about the design and development of this third generation of systems are still not well defined, thus they are focusing an important research activity nowadays (D. Smith and S. Singh, 2006).

This paper introduces the formalism for an intermediate architectural layer aimed at fusing and pro-

M. López-López L., Albusac J., Jesús Castro-Schez J. and Jiménez-Linares L. (2009).
SEMANTICS-PROVIDED ENVIRONMENT VIEWS FOR NORMALITY ANALYSIS-BASED INTELLIGENT SURVEILLANCE.
In *Proceedings of the International Conference on Agents and Artificial Intelligence*, pages 161-166
DOI: 10.5220/0001542701610166

viding with semantics the sensory data obtained from the environment according to the particular necessities of the reasoning components in the upper layers. The fusion of heterogeneous sensory data (D. L. Hall and J. Llinas, 1997) (H. B. Mitchell, 2007) is an important issue to resolve in this respect. This is managed by means of *environment views* which are objects that comprise sensory data which is fused and provided with semantics according to the particular necessities of the reasoning components which require them.

Next, the motivation for introducing an intermediate semantic layer is exposed in section 2. The formal model of the proposed semantic layer is described in section 3. Finally, section 4 concludes the paper.

2 MOTIVATION FOR A SEMANTIC LAYER

Our proposed semantic layer comes to bridge the existing gap between the lower and the upper layers in the general architecture of an intelligent surveillance system. The lower layers are usually responsible for allowing the distribution of sensors throughout the environment and communicating the signals they produce. Such signals are composed of data which is highly device-dependent. On the other hand, the upper layers are usually in charge of inferring a cognitive model of the situation in which the environment is from the captured sensory data. In order to isolate the upper layers from the details of the lower ones, we introduce an intermediate *semantic layer* of which purpose is to process the sensory data, fusing it and providing it with the proper semantics according to the particular necessities of the reasoning components that demand it. Next in this section, we briefly introduce a formal model to detect abnormal behaviours or activities in a monitored environment based on the analysis of the normality of defined concepts (J. Albusac et al., 2008).

When it comes to design an intelligent surveillance system that is able to distinguish between *normal* and *anomalous* situations, three possible approaches are proposed: (1) defining anomalous situations with the collaboration of a human expert; (2) defining normal situations with the collaboration of a human expert; (3) defining normal situations together with the most frequently occurred anomalous ones. According to the first approach, the expert must be able to define every possible anomalous situation that may occur. This turns to be impossible in real scenarios, where the abnormality is unpredictable in most cases. According to the second approach, only the situations considered as normal are defined as these are usually easier to enumerate than the anomalous ones. However, although every non-recognised activity would be marked as anomalous, it would be impossible for the system to establish the degree of potential risk that situation implies. Our work is based on the last approach. In this case, the expert must define not only the set of normal situations, but also the set of the most frequently occurred anomalous ones. Thus, the system is able to recognise every normal and anomalous situation which has been previously defined and the worst case arises when an anomalous situation which has not been previously defined occurs. Similarly to the second approach, in this case the system would not be able to establish the degree of risk that situation implies, however it would be possible to trigger some kind of alarm.

To formalise the normality of a monitored environment, we have elaborated a model in which we define the *problem of surveillance* as the interpretation of a set of perceptions which are provided by a collection of sensors. This model is proposed in (J. Albusac et al., 2008) and briefly included here for convenience.

$$P = \{E_1, E_2, ..., E_n\} \tag{1}$$

Where P denotes the problem of surveillance in a global monitored environment and each E_i denotes a portion of such environment which features are captured by means of a set of sensors. Therefore, each E_i is considered a monitored environment as well.

Also, we can define an environment E_i as a four elements tuple as following.

$$E = < V; O; C; O \times C > \tag{2}$$

where:

- V is the set of variables extracted from the data provided by the sensors. Such extraction task is performed by the semantic layer.
- O is the set of object classes to which the system must pay special attention.
- C is the set of concepts upon which normality analysis processes are performed, such as *normal object trajectories* or *normal object speed*.
- $O \times C$ refers to the correspondence between the sets of objects and that of the concepts. This defines which concepts must be used to analyse the behaviour of each object class, such as *normal pedestrian trajectories* or *normal vehicles speed*.

Next, we define the normality N_{C_i} of some given concept C_i as following:

$$N_{C_i} = < V_i; DDV_i; \Phi_i > \quad (3)$$

where,

- $V_i \subseteq V$ is the subset of variables needed to perform the normality analysis.
- DDV_i is the set of domains of definition for each variable in V_i.
- Φ_i is the set of constraints used to analyse the normality of an object behaviour according to a concept C_i.

The constraints in Φ_i are functions f_{ij} which take a set of variables from V_i as input and return a value within the interval $[0,1]$ as output.

$$f_{ij} : \mathcal{P}(V_i) \longrightarrow [0,1] \quad (4)$$

where 1 implies the maximum degree of constraint satisfaction and 0 the opposite case.

In short, the above definitions mean that to perform the normality analysis of an object behaviour according to a concept C_i, a set of variables V_i extracted from the sensors data V is needed. Besides, some variables in V must be included in several different V_i, which means that such variables are needed to perform several different normality analysis processes. Moreover, according to each normality analysis, each variable in its corresponding V_i has an associated domain of definition DDV_i. Such domains of definition provide each variable in V_i with particular semantics according to the context in which N_{C_i} is performed. Therefore, some given variable may have different semantics according to the different contexts in which it is needed to analyse the normality.

3 PROPOSED SEMANTIC LAYER

The objectives of the proposed semantic layer are:

- To verify and interpret the different data flows provided by the sensors in the environment.
- To fuse the captured sensory data as required (*environment views*) by the normality analysis components.
- To provide the sensory data with semantics according to the specific context in which it is needed.

As shown in Figure 1, the semantic layer is composed of two sub-levels:

- **Interpretation Sub-level.** Which objectives are obtaining, verifying, and interpreting the data flows supplied by the sensors. The output of this sub-level is the data extracted from the sensors and represented in a device-independent common form.
- **Data Fusion and Distribution Sub-level.** Which objective is to provide the components in the reasoning layer with the environment data they require and according to the semantics they demand. That is, to provide each normality analysis component with the environment views they request.

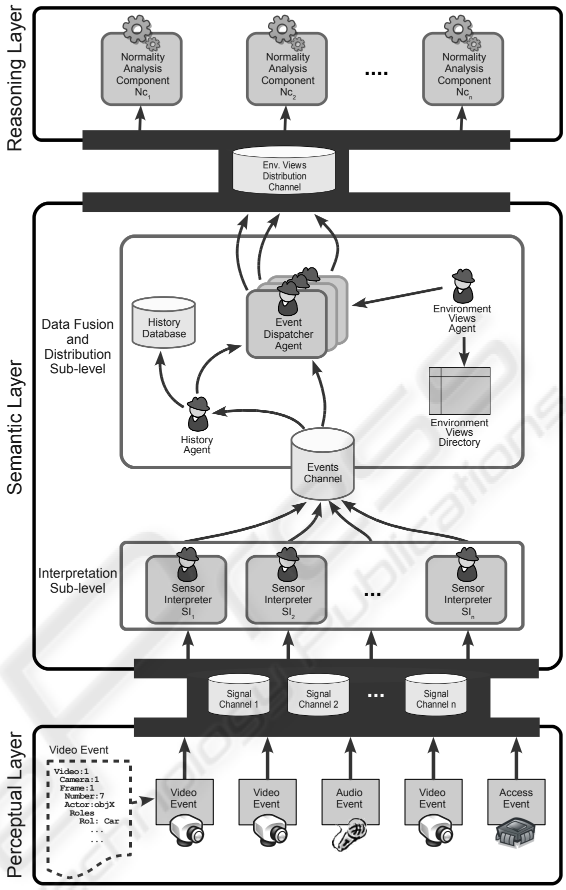

Figure 1: The Semantic layer links the Perceptual and Reasoning layers.

3.1 Interpretation Sub-level

Data provided by each sensor on the environment come to the semantic layer through the *signal channels*. Then, components in the interpretation sub-level are responsible for obtaining, verifying and interpreting such data. These components, named *sensor interpreters*, may be viewed as *agents* which are specialised in interpreting the data provided by some kind sensor.

First, a sensor interpreter must verify and interpret the data provided by the sensor it is specialised in. Such data is highly device-dependent, so we can state that for any sensor s_i installed on the environment, s_i provides strings of data according to some language which generates them.

That is, sensor s_i generates strings of the form:

$$\alpha \in A_{s_i}^* \tag{5}$$

which means that strings α are made up of symbols from some alphabet A_{s_i}. Besides, we can state that for any sensor s_i, there is always a language L_{s_i} associated to it.

$$\forall s_i \; \exists L_{s_i} \text{ such that } L_{s_i} \subseteq A_{s_i}^* \tag{6}$$

Moreover, each language L_{s_i} is described by means of a formal grammar G_{s_i}, defined as:

$$G_{s_i} = (N_{s_i}, T_{s_i}, P_{s_i}, S_{s_i}) \tag{7}$$

where $N_{s_i}, T_{s_i}, P_{s_i}, S_{s_i}$ are the sets of non-terminal symbols, terminal symbols, productions and the start symbol of the grammar G_{s_i}, respectively.

Therefore, we can state that a sensor interpreter is a five elements tuple defined as:

$$SI_{s_i} = < V_{s_i}; SR_{s_i}; G_{s_i}; I_{s_i}; SC_{s_i} > \tag{8}$$

where:

- $V_{s_i} = \{v_{s_i}^1, ..., v_{s_i}^n\}$ is the set of variables extracted from the data provided by sensor s_i.
- $SR_{s_i} = \{sr_{s_i}^1, ..., sr_{s_i}^n\}$ is the set of *semantic rules* defined for the variables extracted from sensor s_i. This is a set of constraints which the incoming data must satisfy in order to determine that the sensor s_i works properly. An example of such a rule might be to check whether some variable's value is within some range of definition.
- G_{s_i} is the grammar that generates the language L_{s_i} that describes the strings of data provided by the sensor s_i (7).
- I_{s_i} is a function which transforms the strings α provided by the sensor s_i into strings β from a common intermediate language IL.

  $$I_{s_i} : \alpha \longrightarrow \beta \tag{9}$$

  such that $\alpha \in L_{s_i}$ and $\beta \in IL$. Besides, such transformation satisfies

  $$m(\alpha) = m(\beta) \tag{10}$$

  where $m(x)$ is a function which returns the meaning of x on the semantic domain.
- SC_{s_i} is the set of signal channels to which the interpreter SI_{s_i} is subscribed to. A signal channel SC_{s_i} is the mechanism by which sensor interpreter SI_{s_i} obtains the data provided by sensor s_i.

Once the interpretation process is finished, sensor interpreters send the extracted data to the *data fusion and distribution sub-level*, through the *events channel*. At this point, data is represented according to an intermediate common language IL to allow the fusion of data extracted from several kinds of sensors and the construction of the environment view objects requested by the normality analysis set of components.

3.2 Data Fusion and Distribution Sub-level

Components in this sub-level are responsible for supplying the normality analysis components in the reasoning layer with the environment data they require and in the form they require it. This lead us to define the concept of *environment view*. An environment view is an object of data consisting of a set of data variables together with associated semantics as required by the normality analysis component that requests it. Each normality analysis component must request an environment view in order to be provided with the data from the environment it needs.

Let $S = \{SI_1, ..., SI_n\}$ be the set of sensor interpreters plugged-in to the system,

$$\forall SI_i \in S, \; \exists V_{SI_i} = \{v_{SI_i}^1, ..., v_{SI_i}^n\} \tag{11}$$

defined V_{SI_i} as the set of data variables provided by sensor interpreter SI_i.

Let V be the set of data variables provided by all sensor interpreters in S,

$$V = \bigcup_{i=1}^{n} V_{SI_i} \tag{12}$$

We define an *environment view* as a two elements tuple as follows:

$$EV_j = < V_j; DDV_j > \tag{13}$$

where:

- $V_j = \{v_j^1, ..., v_j^n\} \subseteq V$ refers to the set of variables included in environment view EV_j.
- $DDV_j = \{DDV_j^1, ..., DDV_j^n\}$ refers to the set of domains of definition for the variables in V_j. These elements specify the proper semantics to their corresponding elements in V_j.

Next, we define how to obtain the variables V_j and the form of the elements DDV_j^i in DDV_j.

Let $V_{jSI_i} \subseteq V_j$ be the subset of variables from V_j provided by sensor interpreter SI_i, named *component view* of EV_j.

$$V_j = \bigcup_{i=1}^{n} V_{jSI_i} \text{ and } V_{jSI_i} \subseteq V_{SI_i} \tag{14}$$

which means that the set of variables V_j of an environment view EV_j is the aggregation of the component views V_{jSI_i} extracted from the sets of variables V_{s_i} provided by each sensor interpreter SI_{s_i}. Of course, the component view V_{jSI_i} for a given sensor interpreter SI_{s_i} may be the empty set, which means that the environment view EV_j does not require any data variable from sensor interpreter SI_{s_i}.

Moreover, data provided by sensor interpreters may be given as crisp or vague values. For example, a presence detection sensor interpreter may return a crisp variable value indicating *presence detected*, which may be given as a boolean *true* value. Also, a video camera may return some linguistic value indicating that a *person moving fast* has been identified in a monitored area. Thus, we need a common data type which allows us to represent every possible kind of value provided by any sensor interpreter in *S*. For that purpose, we propose the use of *trapezoidal functions*, as defined in (15) (J. J. Castro-Schez et al., 2004b).

$$\prod(u;a,b,c,d) = \begin{cases} 0 & u < a \\ \frac{(u-a)}{(b-a)} & a \leq u < b \\ 1 & b \leq u \leq c \\ \frac{(d-u)}{(d-c)} & c < u \leq d \\ 0 & u > d \end{cases} \tag{15}$$

The reason for choosing this function is that it is a suitable way to represent both, categorical and numerical values and in both cases such values may be given to the user by means of linguistic terms, which is convenient in order to provide them with information expressed in their own terminology. Besides, the inference rules in the normality analysis components can be expressed using linguistic variables and fuzzy theory techniques for dealing with the inherent uncertainty of the environment data (J. J. Castro-Schez et al., 2004a).

When designing a new normality analysis component, a new environment view must be also specified indicating its requirements in terms of environment data and its associated semantics.

The main components in this sub-level are the *event dispatcher agents* (Figure 1), which main functions are: (1) collecting the data required by the normality analysis component; (2) processing such data according to the requirements of the normality analysis components to be served; and (3) distributing the data to those normality analysis components that demand it.

The event dispatcher agents are the components responsible for distributing the sensory data to the normality analysis components that require it in form of environment views. This process is as follows:

1. An event dispatcher agent keeps waiting until some event of data is triggered.
2. When a new environment event is triggered, some event dispatcher agent collects the incoming data and asks the *environment views agent* for the normality analysis components which requested an environment view containing the obtained data.
3. The selected items (environment views) may need to be completed with previously obtained data from the *history database*, because the incoming data may be insufficient to construct a complete environment view.
4. Once an event dispatcher agent has obtained the complete chunk of data which make up an environment view, it processes it according to DDV_j in order to provide data with the requested semantics.
5. Finally, the event dispatcher agent transfers the chunk of data to the normality analysis component (or components) which requested the environment view object just constructed.

The environment views are stored in the *environment views directory*. This component is managed by the *environment views agent*, which is responsible for both publishing new data in the directory and retrieve existing data under request of an event dispatcher agent. When a new normality analysis component is plugged-in to the system, a new environment view must be defined and published in the environment views directory. We define the structure of this component as a collection of *entry* elements. An entry element is a two elements tuple composed by (N_{C_i},EV_j) elements. The element N_{C_i} refers to the identifier of some normality analysis component, whereas the EV_j element refers to a defined environment view. This way, some environment view may be associated with several normality analysis components when needed.

When the data contained in a new environment event is insufficient to construct an environment view object, the event dispatcher agent responsible for attending that event must go somewhere else to obtain the missing data. Such place is the *history database*. This component is managed by the *history agent* which responsibilities are both to store data incoming from the event channel in the history database and to retrieve existing data from the database under request of an event dispatcher agent. The history database is composed of a collection of *entry* elements. Each entry consists in a three elements tuple (SI_i, V_{s_i}, Moment), such that SI_i is the identifier of the sensor interpreter which provided the data, V_{s_i} is the set of variables extracted from sensor interpreter SI_i, and Moment is the time and date when such information was retrieved by the system.

Once an event dispatcher agent has completed the chunk of data that comprises an environment view, it sends a message containing such data through the *environment views distribution channel*. This channel supports a broadcast-like communication mode, such that every normality analysis component must subscribe to it and they can just receive those messages that are sent to them. This approach has as its main advantage the fact that only one communication point is needed to be known by both, the event dispatcher agents and the normality analysis components.

4 CONCLUSIONS

The trend in the design and development of intelligent surveillance systems is to use not only the visual information provided by a set of video cameras, but also to use other kinds of sensors to allow the system to handle a more accurate knowledge of the monitored environment. In this respect, the fusion of sensory data plays an essential role as different sensors provide data in a variety of different forms. Besides, the intelligent surveillance based on normality analysis requires that the same sensory data can is used in different analysis contexts with different semantics. We have presented an architectural layer to fuse and provide with semantics sensory data according to the particular requirements of the normality analysis components which require it. We have defined the concept of environment view as an object that contains the data requested for a normality analysis component. Besides, the proposed architecture allows for the easy scalability in terms of both, the sensors installed in the environment and the normality analysis components plugged-in to the system. Adding a new kind of sensor entails designing a new sensor interpreter capable of interpreting the data sent by such kind of sensor and leaving it available to the rest of components of the system. On the other hand, adding a new normality analysis component entails to define a new environment view, according to the requirements of data and semantics of the newly added component.

ACKNOWLEDGEMENTS

This work has been founded by the Regional Government of Castilla-La Mancha under Research Project e-PACTOS (ref. PAC-06-141), Sarasvati (ref. PBC06-0064) and the Spanish Ministry of Education and Science under the Research Project TIN2007-62568.

REFERENCES

D. L. Hall and J. Llinas (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23.

D. Smith and S. Singh (2006). Approaches to Multisensor Data Fusion in Target Tracking: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 18(12):1696–1710.

H. B. Mitchell (2007). *Multi-Sensor Data Fusion: An Introduction*. Springer-Verlag.

J. Albusac, D. Vallejo, L. Jimenez, J. J. Castro-Schez, and L. Rodriguez (2008). Intelligent Surveillance based on Normality Analysis to Detect Abnormal Behaviors. *Submitted to International Journal of Pattern Recognition and Artificial Intelligence*.

J. J. Castro-Schez, J. L. Castro, and J. M. Zurita (2004a). Fuzzy repertory table: a method for acquiring knowledge about input variables to machine learning algorithm. *IEEE Transactions on Fuzzy Systems*, 12(1):123–139.

J. J. Castro-Schez, N. R. Jennings, X. Luo, and N. Shadbolt (2004b). Acquiring domain knowledge for negotiating agents: a case of study. *International Journal of Human-Computer Studies*, 61(1):3–31.

L. Rodriguez-Benitez, J. Moreno-Garcia, and J. J. Castro-Schez (2008). Automatic Object Behaviour Recognition from Compressed Video Domain. *Image Vision and Computing*. doi: 10.1016/j.imavis.2008.07.002.

M. Valera and S. A. Velastin (2005). Intelligent Distributed Surveillance Systems: A Review. *IEEE Proceedings on Vision, Image, and Signal Processing*, 152(2):192–204.

P. Remagnino, A. I. Shihab, and G. A. Jones (2004). Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition*, 37(4):675–689.

R. C. Luo, C. Yih, and K. L. Su (2002). Multisensor Fusion and Integration: Approaches, Applications, and Future Research Directions. *IEEE Sensors Journal*, 2(2):107–119.

R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson (2000). A System for Video Surveillance and Monitoring. Technical report, The Robotics Institute, Carnegie Mellon University.

W. Hu, T. Tan, and L. Wang (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352.