Object-Event Graph Matching for Complex Activity Recognition

Alexander Bauer and Yvonne Fischer

***Abstract*—In security, the most relevant criminal and terrorist activities are often of high complexity: they involve several entities interacting sequentially and simultaneously over an extended time interval. In this paper, we present a powerful approach for complex activity recognition and analysis using graph representation and matching. It is based on the representation of activities in terms of objects, events and processes, which are modeled as nodes of an attributed relational graph (ARG). The recognition of complex activities, taking into account observation uncertainty and incompleteness, is performed using graph matching of template graphs and the data graph. The data graph represents observations of objects, events and processes collected from low-level signal processing and other information sources. The models of the complex activities to be detected are represented as template graphs. Markov chain Monte Carlo sampling is proposed to infer probabilities of activity occurrence, object involvement and event occurrence for detection, event prediction and sensor management in complex activity recognition problems. The suggested method is illustrated using a toy example from maritime surveillance.**

***Index Terms*—situation awareness, surveillance, activity recognition, graph matching, markov chain monte carlo, sensor management**

I. INTRODUCTION

For a decision maker in a surveillance control center, it is important to have excellent situation awareness. Due to the advances in sensor networks and digital signal processing, surveillance technology of today becomes more and more capable to support the decision maker by automated detection, tracking and identification of relevant elements of the environment (such as tracking of people in video streams). However, as pointed out by Endsley's definition of situation awareness (SA), it is not enough to be aware of all relevant elements of the environment, but also to comprehend their meaning and to project their status in the future [1]. In most surveillance applications, this is equivalent to the recognition and analysis of activities performed by humans or human-controlled vehicles (ships, cars, etc.). Accordingly, activity recognition and behavior analysis have been identified as the technical solution to support higher levels of SA [2].

Still, generally applicable and reliable methods for activity recognition are missing which can address the major challenges:

1. Activities, especially in the security domain, have high complexity: they involve several objects interacting sequentially over an extend time interval, in order to achieve their goal.
2. Even advanced surveillance systems will hardly be capable to completely observe all objects and interactions involved in a complex activity – however for the decision maker, it is important to detect such activities before they are completed in order to anticipate them.
3. Most activities of interest in security applications do not occur frequently enough, such that machine learning methods can be efficiently applied. Solutions for the recognition of such activities must therefore be able to tap on expert knowledge but should be at the same time open to learning methods to improve their performance over time.

To advance towards a general solution for these challenging requirements, we present the definition of object-event graphs, a graph representation for the recognition and analysis of complex activities involving multiple objects and interactions. They provide a powerful framework to define activities at a level of abstraction which is suitable for expert knowledge representation. The formulation of the activity recognition problem as an inexact maximum common sub-graph matching of object-event graphs is able to deal with uncertainty and incompleteness in the observation data and the investigation of unmatched nodes can give valuable hints for sensor cueing and event prediction.

II. RELATED WORK

Several methods originating from artificial intelligence research have been investigated for activity recognition. Hidden Markov models (HMM) have been successfully applied to single object behavior classification, for example in aircraft mission classification [3]. However, the underlying Markov assumption is violated in more complex behavior

Manuscript received October 15, 2010. This work was supported by the European Commission as part of the project WiMAAS.

Alexander Bauer is with the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, 76131 Karlsruhe, Germany.

Yvonne Fischer is with the Institute for Anthropomatics - Vision and Fusion Laboratory (IES) at Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany. (e-mail: yvonne.fischer@kit.edu).

patterns involving long-term temporal correlation, and extensions of the standard HMMs or the use of dynamic Bayesian networks (DBN) have been proposed to tackle these problems [4]-[5]. Still, these methods are limited to single objects and are strongly dependent on training data, as the HMM states do not carry semantic meaning which is necessary for expert knowledge representation. Hybrid approaches based on predicate logic [6]-[8], constraint satisfaction [9] and grammar parsing [10]-[11], have been applied to perform recognition of complex activities based on event primitives extracted from low-level classifiers. However their abilities to treat uncertainty are often limited and are not built on a consistent probabilistic representation. The treatment of incomplete observation and the cueing of sensors to benefit from additional observations in order to reject or confirm an activity hypothesis have been mostly neglected in previous approaches.

Graph matching methods have been widely applied in image interpretation and pattern recognition for several decades [12]. Especially the use of attributed relational graphs provides an excellent framework for the representation and recognition of relational structures in images [13]. Complex activities exhibit those strong relational dependencies as well: multiple objects involved in different interactions occur in specific temporal arrangements. The potential of graph representations for high level fusion has been recognized and provided the motivation for the development of optimization-based graph matching methods for forensic applications [14]. Due to the computational complexity of graph matching methods, they have only recently been proposed for the use in online activity recognition [15]. To further promote their use for online applications, we introduce a generic graph based representation for multi-object activities and highlight how online activity recognition and analysis could be efficiently performed by maximum common subgraph matching using Monte Carlo sampling methods.

III. OBJECT EVENT GRAPHS FOR ACTIVITY MODELING

A. Idea

Pattern recognition using graph matching is based on the representation of template pattern and data pattern as attributed relational graphs. An attributed relational graph (ARG) is a directed graph that is defined by the tuple $G=(V,E,A_V,A_E)$ where V is a set of nodes, E is a set of edges, A_V is a set of node attributes and A_E is a set of edge attributes.

We exploit the pattern representation flexibility of ARGs by combining objects and events to create a new representation of complex activities which we call object-event graphs. They allow the modeling of activities involving complex multi-object interactions and temporal relations. The template graph represents all objects and events involved in the activity, whereas the data graph contains all observed objects and events detected in a relevant time frame respectively. Based on this representation, maximum common subgraph matching enables the recognition of activities in uncertain and incomplete observation data and even prior to the completion of the activity, as only a subset of all nodes in the template graph has to be present in the data graph. Unmatched nodes and edges in the template graph can be further exploited to direct the attention of a human operator or a sensor system to undetected events, undetected objects or events happening in the future. Figure 1 illustrates the relation between template graph, data graph, the node match associations and unmatched nodes. Objects, processes and events appear as different types of nodes in the object-event graph.

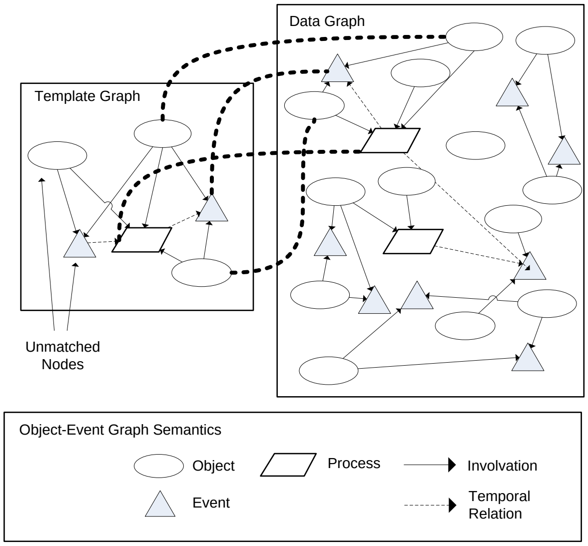

Fig. 1. Illustration of the result of a maximum common subgraph matching between a template object-event graph and a data object-event graph. Unmatched nodes in the template graph give strong indications on undetected objects and prospective events.

B. Node Types

Nodes in the object-event graph describe physical and non-physical entities necessary to represent an activity:

- Object nodes $v_O \in V_O$ represent physical objects involved in an activity. As an example, they can be humans or any object they interact with. The attributes of object nodes describe the characteristic features of the object involved in the represented activity (e.g. type, dimensions, appearance, identity etc.).
- Event nodes $v_E \in V_E$ describe relevant state changes of objects and object interactions which do not have a significant time extent (e.g. a car stops). The attributes of event nodes describe the characteristic features of the event (e.g. type, time of occurrence, etc.)
- Process nodes $v_P \in V_P$ describe state changes and object interactions which take place over a significant time extent (e.g. an object is traveling from one place to another). Their attributes describe the features of the process (e.g. type of process, start time, end time etc.)

C. *Edge Types*

Relations between objects, events and processes are established using the following edge types:

- Involvement edges $e_I \in E_I$ describe the involvement of an object in an event or a process. Their attributes describe attributes of the involvement, for example the role of the object in the process or event.
- Temporal edges $e_T \in E_T$ describe the temporal relations between events and processes (e.g. before, after, while, etc.) An excellent definition of temporal relations has been developed by Allen [16].

D. *Template and Data Graph*

Using a template graph $G^T = (V^T, E^T, A_V^T, A_E^T)$, an expert models the activities of interest in terms of object, process and event primitives and establishes the corresponding involvement and temporal edges. Expected variations in event or object attribute values can be modeled using probability distributions. For temporal edges, intervals (e.g. “2-5h after”) are a suitable representation for expert modeling, and can be expressed as uniform distribution.

The data graph $G^D = (V^D, E^D, A_V^D, A_E^D)$ is established and maintained from observation data. In general, the data graph will be much larger than the template graph, as in most applications there are a lot of objects and events in the observed area which are not involved in the activity of interest. The generation and update mechanism for the data graph depends on the domain of application and the means of observation. We suggest to extract it from a consistent object-oriented representation of objects and their attributes, as the object-oriented world model (OOWM) suggested in [17], established by fusion of multiple source such as sensors, human observers, or other information sources. Events and processes are either directly provided by the information sources or they are derived from the object-oriented representation using trajectory analysis or from object attribute changes.

E. *Template Graph Example*

To illustrate the object-event graph representation for complex activities, we choose an illegal immigration activity in the Mediterranean Sea: A small wooden boat departs from the Tunisian coast, approaches the island Lampedusa and finally lands at the beach to unload the immigrants. This natural description of an illegal immigration activity is translated into an object-event graph representation. Fig. 2 shows the complete template graph. For each node and edge in the graph, attributes are added which will constraint the matching with data nodes in the matching process.

The example shows how easily an expert can express, discuss and revise an activity representation using object-event graphs, paying full respect to the all relevant objects, interactions and temporal relations involved in the activity. To improve the matching performance, the representation of attribute value distributions can be learned and improved from training data.

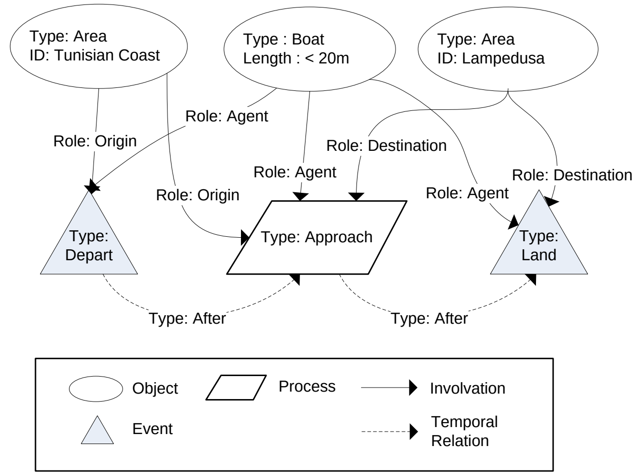

Fig. 2. Example for a template graph modeling an illegal immigration activity in the Mediterranean Sea.

IV. OBJECT-EVENT GRAPH MATCHING

To perform the matching between template and data graph, we propose the following probabilistic formulation of the inexact subgraph matching problem.:

Let Ω be a set of subgraph matches of G^T and G^D such that for $\omega \in \Omega$,

1) $\omega = (V_M^T, E_M^T, f_V, f_E)$;
2) $V_M^T \subseteq V^T$ is a set of matched template nodes;
3) $E_M^T \subseteq E^T$ is a set of matched template edges;
4) $f_V : V_M^T \rightarrow V^D$ is an injective node mapping;
5) $f_E : E_M^T \rightarrow E^D$ is an injective edge mapping.

The measure of quality for a match is expressed as the posterior probability of ω:

$$\begin{aligned} P(\omega|A_E^T, A_V^T, A_E^D, A_V^D) &\propto P(\omega) \\ \times P(f_V|A_V^T, A_V^D) &P(f_E|A_E^T, A_E^D) \end{aligned} \tag{1}$$

The prior of a match is assumed to be composed of two binomial distribution with unmatched node probability p_V and unmatched edge probability p_E:

$$\begin{aligned} P(\omega) \propto & \binom{|V^T|}{|V_M^T|} (1-p_V)^{|V_M^T|}\, p_V^{|V^T|-|V_M^T|} \\ \times & \binom{|E^T|}{|E_M^T|} (1-p_E)^{|E_M^T|}\, p_E^{|E^T|-|E_M^T|} \end{aligned} \tag{2}$$

For the likelihood of a match, we suggest to assume independence between individual edge match $\langle e^D, e^T \rangle$ and node match $\langle v^D, v^T \rangle$ probabilities:

$$\begin{aligned} P(f_V|A_V^T, A_V^D) &= \textstyle\prod_{v \in V_M^T} P\big(\langle v, f_V(v)\rangle | A_V^T, A_V^D\big) \\ P(f_E|A_E^T, A_E^D) &= \textstyle\prod_{e \in E_M^T} P\big(\langle e, f_E(e)\rangle | A_E^T, A_E^D\big) \end{aligned} \tag{3}$$

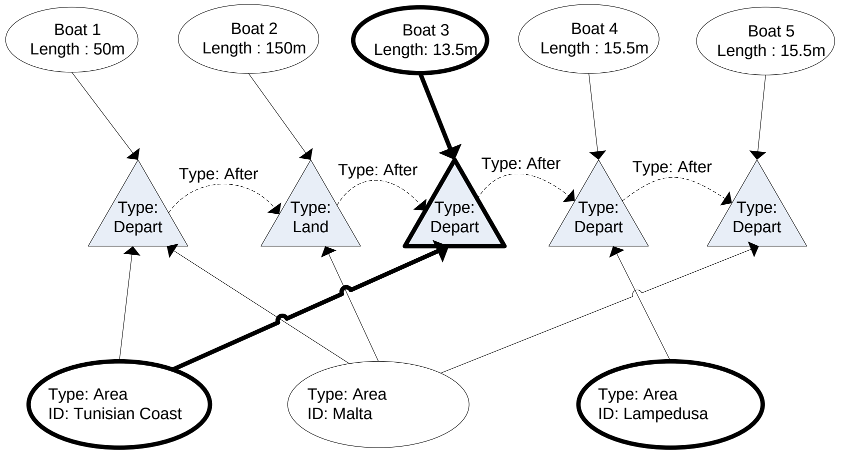

Fig. 3. Example for a data graph representing ship traffic around Lampedusa, Malta and the Tunisian Coast. Nodes in the MAP match are highlighted by a bold border. For clarity, temporal relations between events are only shown for direct ancestor/successors event nodes.

The problem of inexact maximum common subgraph matching is then formulated as the maximum a posteriori (MAP) estimate of ω:

$$\overline{\omega} = \arg\max_{\omega} P(\omega | A_E^T, A_V^T, A_E^D, A_V^D) \tag{4}$$

Finding $\overline{\omega}$ is a typical discrete optimization problem. From $\overline{\omega}$, one way to decide if the activity has been recognized would be to count the number of matching nodes and edges, and to derive a suitable decision function from it.

By sampling from the posterior distribution, it is also possible to estimate the probability that at least half of the template nodes have been matched. To derive the probability of such a subset $X \in \Omega$, an indicator function I_X is defined for that set and the expected value is calculated:

$$P_X = \mathbb{E}_{\omega \in \Omega}\{\mathbf{I}_X(\omega | A_E^T, A_V^T, A_E^D, A_V^D)\} \tag{5}$$

By defining other indicator functions, sophisticated analysis can be performed:

- *Determining suspicious objects* based on the probability that a specific object node v_O^D in the data graph has been matched to a template node (*activity involvement probability*),
- *Inferring undetected events* from the probability that a specific event node in the template graph has not been matched,
- *Inferring undetected objects* from on the probability that a specific event node in the template graph has not been matched.

A similar approach has been successfully used for aerial image interpretation of industrial sites to infer undetected objects in the image [18]. Sampling from posterior probability distributions of large probability spaces can be efficiently handled using Markov Chain Monte Carlo (MCMC) methods. Lee et al have recently shown that these methods can outperform traditional methods in graph matching problems [19].

V. EXAMPLE

To illustrate the behavior of the proposed algorithm, the following toy example is presented. We assume the scenario that ship traffic is observed by sensors around the island Lampedusa, the island Malta and the Tunisian Coast, and objects and events are detected in the following chronological order: Boat 1, a medium size boat departs from the Tunisian Coast. Boat 2, a large tanker, lands in Malta. Boat 3, a small boat, departs from the Tunisian coast. Boat 4, a small boat, departs from Lampedusa and boat 5, a small boat, departs from Malta.

As parameters for the unmatched node and unmatched edge probability (see Section IV) we choose $p_V = 0.9, p_E = 0.999$. The probability of match between data and template nodes is set to 0.999 if the attributes agree and 0.001 otherwise. The same probabilities apply for the matching of edges. We use MCMC to draw 10^5 samples from the posterior distribution (1) based on the node association dynamics proposed in [19]. Fig. 3 shows the data graph of our example and the match result. The nodes and edges belonging to the match $\overline{\omega}$ with maximum a posteriori probability are highlighted by a bold border/line.

Besides the determination of the MAP estimate, we also estimate the *activity involvement probability* for each object node from the samples (see Table 1). The estimate for the probability that half of the nodes in the template graph are matched with the data graph is 20%.

TABLE I
POSTERIOR ACTIVITY INVOLVEMENT PROBABILITY

Node name	Activity involvement probability
Boat 1	<0.1 %
Boat 2	<0.1 %
Boat 3	42 %
Boat 4	14 %
Boat 5	14 %

VI. DISCUSSION

The results show that the algorithm generates plausible results: Boat 3, receiving 42% involvement probability, is the only boat that departs from the Tunisian Coast and its length corresponds to the boat modeled by the template graph. Boat 4 and 5 also match the attributes of the boat in the template graph, however they depart from other places and therefore receive a lower probability of involvement (14%). The length of the other boats does not match with the template graph object, and therefore they receive negligible activity involvement probability. This demonstrates that the proposed formulation of the posterior probability takes account for both the structure of the object-event graph and the attributes of the nodes.

As the MAP match in this example only spans a subset of the template graph, further observation of the activity is needed to confirm the occurrence of an illegal immigration activity. From the missing nodes in the template graphs, clues for the tasking of sensors can be derived. For example, from the template graph, unmatched events succeeding the events that are part of the MAP match suggest further sensor employment. In the example, investigating if boat 3 is approaching Lampedusa after its departure could be a useful sensor management action.

As the proposed method makes use of the structural relations between objects, events and processes, it is of course sensitive to structural variations of the activity and might not be able to detect them if strong variations occur. This could be overcome by formulating the possible variations by a graph grammar that generates a set of template graphs to match with the data graph. If only temporal order of events is an issue, reducing the edge match probability p_E for temporal edges can increase the robustness concerning variations in event order.

VII. CONCLUSION

The recognition of complex activities and supporting it through dynamic management of sensors is a challenging problem. Uncertain and incomplete observations, scarce training data and the need for early detection are challenges which cannot be tackled with traditional pattern recognition methods. Object-event graphs as a new representation for complex activities, in combination with advanced inexact subgraph matching methods, exhibit very promising characteristics to attack these challenges: They support expert knowledge representation, uncertainty modeling as well as incomplete matching, providing useful hints for the dynamic management of sensors. Next, their performance will be evaluated for applications in video and signal-based surveillance in maritime surveillance (based on object-level data fusion, as described in [20]), land-border and building security applications.

ACKNOWLEDGMENT

We would like to thank our colleagues Robert Gering and Daniel Flöter who implemented a graph editing and visualization environment for the graphs and the match results. We also would like to thank the developers of the Java Universal Network/Graph Framework (JUNG) [21], an open-source software library for graph representation and visualization.

REFERENCES

[1] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems: Situation awareness," *Human factors*, vol. 37, no. 1, pp. 32-64, 1995.

[2] J. J. Salerno, "Where's level 2/3 fusion - a look back over the past 10 years," in *2007 10th International Conference on Information Fusion*. IEEE, July 2007.

[3] M. Andersson, "Classification of aerial missions using hidden markov models," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Lecture Notes in Computer Science, vol. 2711, pp. 125-136, 2004.

[4] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 838-845, July 2005.

[5] S. Hongeng, R. Nevatia, and F. Bremond, "Video-based event recognition: activity representation and probabilistic recognition methods," *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 129-162, November 2004.

[6] J. Ijsselmuiden and R. Stiefelhagen, "Towards high-level human activity recognition through computer vision and temporal logic," in *KI 2010: Advances in Artificial Intelligence*, Lecture Notes in Computer Science, vol. 6359, pp. 426-435, 2010.

[7] M. S. Ryoo and J. K. Aggarwal, "Semantic representation and recognition of continued and recursive human activities," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 1-24, November 2008.

[8] J. Gonzàlez, D. Rowe, J. Varona, and F. Xavier Roca, "Understanding dynamic scenes based on human sequence evaluation," *Image and Vision Computing*, vol. 27, no. 10, pp. 1433-1444, September 2009.

[9] V. T. Vu, F. Bremond, and M. Thonnat, "Automatic video interpretation: a novel algorithm for temporal scenario recognition," *IJCAI'03: Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pp. 1295-1300, 2003.

[10] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 852-872, August 2000.

[11] A. Bauer and Y. Fischer, "Task-oriented situation recognition," in *Proceedings of SPIE*, vol. 7709, 2010.

[12] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, pp. 265–298, 2004.

[13] R. Cesar, E. Bengoetxea, I. Bloch, and P. Larranaga, "Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms," *Pattern Recognition*, vol. 38, no. 11, pp. 2099-2113, November 2005.

[14] K. Sambhoos, R. Nagi, M. Sudit, and A. Stotz, "Enhancements to high level data fusion using graph matching and state space search," *Information Fusion*, December 2009.

[15] G. Levchuk, A. Bobick, and E. Jones, "Activity and function recognition for moving and static objects in urban environments from wide-area persistent surveillance inputs," Proceedings of SPIE, vol. 7704, 77040P, 2010.

[16] J. F. Allen and G. Ferguson, "Actions and events in interval temporal logic," *Journal of Logic and Computation*, vol. 4, pp. 531-579, 1994

[17] A. Bauer, T. Emter, H. Vagts, and J. Beyerer, „Object oriented world model for surveillance systems", *In Future Security: 4th Security Research Conference*, pp. 339–345, Fraunhofer Press, 2009.

[18] A. Bauer, "Probabilistic scene models for image interpretation," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, Communications in Computer and Information Science, vol. 81, pp. 562-571, 2010.

[19] J. Lee, M. Cho and K. M. Lee, "A graph matching algorithm using data-driven markov chain monte carlo sampling," *ICPR'10: International Conference on Pattern Recognition*, IEEE, 2010, in press.

[20] Y. Fischer and A. Bauer, „Object-oriented sensor data fusion for wide maritime surveillance", Water side security conference 2010, in press.

[21] http://jung.sourceforge.net/