Interpretation of Human Motion in Image Sequences using Situation Graph Trees

Pau Baiget*, Jordi Gonzàlez+, Javier Orozco*, Xavier Roca*

* *Centre de Visió per Computador, Campus UAB, Edifici 0, 08193 Bellaterra, Spain*
pbaiget@cvc.uab.es

+ *Institut de Robòtica i Informàtica Industrial (UPC-CSIC)*
Edifici U, Llorens i Artigas 4-6, Barcelona 08028, Spain

Abstract Evaluation of human behaviour patterns in determined scenes has been a problem studied in social and cognitive sciences, but now it is raised as a challenging approach to computer science due to the complexity of data extraction and its analysis. Results obtained in this research will be helpful in cognitive sciences, above all in the human-computer interaction and the video-surveillance domain. Our information source is an image sequence previously processed with pattern recognition algorithms, to extract quantitative data of the trajectories performed by the agents within the scene. Reasoning about human behavior makes necessary the inclusion of machine learning techniques, in order to represent those behaviours in a qualitative manner, allowing natural language explanation of the scene. This is achieved by means of a rule-based inference engine called F-Limette and a behaviour modelling tool based on Situation Graph Trees. The success of this approach depends on the precision of the image analysis system, the selection of suitable reasoning tools and the design of useful behaviour models. The model was tested in a street scene and the agents of interest were pedestrians. Textual descritpions are generated which qualitatively describe the observed behavior. Experimental results are provided by defining three different behaviors in a pedestrian crossing. This will allow us to confront sociological theories about human behaviour, whose quantitative base is at present being computed from statistics and not from semantic concepts.

Keywords: Motion Analysis and Recognition, Pattern Recognition, Human Behavior Analysis, Cognitive Vision.

1 Introduction

Any system trying to model human behavior has to deal with the uncertainty due to the *semantic gap* [12]. *Uncertainty* arises because of the impossibility of modelling all the possible human behaviors in a given discourse domain. Therefore, uncertainty has to be considered, so logic predicates (whose value is *true* or *false*) must change to fuzzy predicates, with an associated numerical value of truth. Thus, the semantic gap refers to the conceptual ambiguity between the image sequence and its possible interpretations. For example, an agent walking in a parking lot moving his face back and forth can be interpretated as a person searching his parked car or a thief deciding which car to steal.

Integration refers to the interpretation process from quantitative to qualitative knowledge, and should deal with the semantic gap between the quantitative information obtained from pattern analysis procedures, and the conceptual description used for reasoning [5]. As a result of integration, numerical information obtained from image sequences can be used to instantiate qualitative predicates and, the other way round, conceptual knowledge are used to assist pattern analysis processes [13].

In this work, we assume that the geometrical information is provided uninterruptedly over time. Thus, plausible situations will be organized into a temporal and conceptual hierarchy to generate behaviour descriptions.

The main objective of this paper is to demonstrate the flexibility of the modelling tool called *Situation Graph Trees* [1], which has been used in traffic videosequences domain [6], by considering the hu-

man behavior interpretation domain.

This paper is structured as follows: next section reviews existing approaches. Section 3 shows the integration process which transforms quantitative data to qualitative predicates. In Section 4 we show an example of Situation Graph Tree describing the expected human behavior in a roadway scene and we describe the framework involved in textual description generation. Finally, Section 5 concludes the paper and shows future lines of research.

2 Related work

A survey of systems creating high-level descriptions from image sequences is presented in [3]: several procedures can be chosen to develop a motion understanding system, depending on the requirements of the discourse domain. Motion understanding systems rely on knowledge based on the geometry of the scene and the numerical data obtained from pattern analysis procedures.

As said before, semantic interpretation may lead to uncertainty, due to the vagueness of the semantic concepts utilized, and the incompleteness, errors and noise in the agent state's parameters. Therefore, uncertainty prevents of categorizing in a precise way the integration of the agent state. In order to cope with this issue, integration can be learnt using a probabilistic framework: PCA and Mixtures of Gaussians [11], Belief Networks [9, 14] and Hidden Markov Models [2, 4] provide examples. However, in certain multi-agent domains, the representation of all the possible behaviours using these models may become a quite complex process, and the estimation of the transition probabilities would be unreliable without a large amount of training data. Alternatively, Fuzzy Metric Temporal Logic (FMTL) also copes with the temporal and uncertainty aspects of integration in a goal-oriented manner [16]. This predicate logic language treats dynamic occurrences, uncertainties of the state estimation process, and intrinsic vagueness of conceptual terms in a unified manner. The main advantage of FMTL over the previously referred algorithms relies on the promise to support not only the execution, but also the diagnosis steps during the continuous development and test of ISE systems [7].

Once the uncertainty is modelled using either a mathematical or logical formalism, semantic predicates are classified according to several criteria, such as the specialization relationship [10], the semantic nature [14] or the temporal order [9] of such predicates. In essence, a suitable behaviour model should explicitly represent and combine the specialization, temporal, and semantic relationships of its constituent conceptual predicates. In this paper, this is accomplished by the Situation Graph Trees modelling tool.

3 Integration: from tracking to understanding

The generation of textual descriptions from image sequences is an important requirement for analyzing the results of human behavior interpretation systems. Based on the geometrical information obtained in tracking processes, logic predicates are instantiated by employing FMTL. As a result, temporally and conceptually isolated logic statements are instantiated for each frame, and embedded into a temporal and conceptual context by using SGTs [1].

For this purpose, the concept of *generically_describable_situation* presented in [12] is used: a situation consists of an agent state, together with the potential reactions that such an agent can perform in such a state. SGTs organize the set of plausible situations into a temporal and conceptual hierarchy. Thus, on the one hand, SGTs represent the temporal evolution of situations, and a set of potential situations are specified for each situation. That means, predicate evaluation is performed in a goal-oriented manner: given a situation, only its successors will be evaluated in the next time step. On the other hand, each situation can be described in a conceptually more detailed way, thus allowing to establish conceptual descriptions with a certain level of abstraction and specificity. Consequently, Situation Graph Trees provide a deterministic formalism to represent the knowledge required for human behavior evaluation, where *behavior* refers to human agent trajectories which acquire a meaning in an specific scene.

3.1 Semantic predicates for human behavior analysis

A set of semantic features is first established, which is derived from the numerical state of the agent. Suit-

able features are determined by the nature of the human agent state's parameters which may refer to dynamical, positional and postural properties of the actor. For example, motion verbs, such as accelerating could be instantiated by evaluating the history of the spatial and speed parameters of the agent state.

In our experiments, the quantitative description of the state of the agent is obtained through a segmentation process based on Horprasert algorithm [8] and on a subsequently state estimation method [15] for each time step. Consequently, the following information is provided for each time step:

1. The 2-D spatial position *pos* of the agent *Agent* in the floor plane, in addition to the velocity *vel* and the orientation *or*. These three parameters are called the *spatial status* of the agent, which is generic for Image Sequence Evaluation systems. This information is commonly obtained using visual tracking procedures.

2. The action label *aLabel* which the actor is performing, obtained with respect to the velocity *vel*. Thus, we may differentiate between *walking*, *standing* or *running*.

All this knowledge is comprised in the following attribute scheme, or *state vector of the agent*:

$$has\_status(Agent, pos, or, vel, aLabel).$$

In order to model the behavior of human agents, we first define an a-priori terminology which consists of logic predicates related to the information to be described about the scene. Therefore, the terminology actually restricts the discourse domain, and consists of logic-rules and facts regarding the state of the agent, its relationship with the environment, and information about the context. Note that the problem domain itself provides the knowledge required to design such a terminology.

Based on the quantitative information of the state vector of the agent obtained from tracking procedures, the aim is to to associate conceptual interpretations for such numerical data. For this purpose, we next address three different strategies to accomplish such an abstraction process, according to the source of knowledge which is exploited to generate the qualitative description.

3.2 About the spatial information of an actor within a scene

Quantitative state parameters are associated to concepts like *moving*, *small*, *left*, or *briefly* with a fuzzy degree of validity characterizing how good a concept matches the numerical quantity. As a result, the speed and orientation parameters of the state vector will be associated to fuzzy attributes, thus allowing the instantiation of logic predicates such as:

$$has\_speed(Agent, Value),$$
$$has\_direction(Agent, Value).$$

3.3 About the relationship of an actor with respect to its environment

Spatial relations are derived by considering the positions of the agents and other static objects in the scene. In this case, a conceptual scene model is required to describe the spatial coordinates of the agent with respect to static objects, other agents, and specific locations within the scene. This description is implemented by applying a distance function between the positions of different agents/objects in the scene. Subsequently, a discretization of the resulting distance value is obtained by using Fuzzy Logic:

$$is\_alone(Agent, Proximity),$$
$$has\_distance(Agent, Patiens, Value).$$

Also, predicates concerning the spatial properties of the agent with respect to static objects (predefined at the conceptual scene model) can be instantiated. For example, the predicate:

$$on\_waiting\_line(Agent, WLine),$$

checks whether the spatial position of the agent *Agent* in the scene is inside the sideway segment defined as *WLine*.

3.4 About the action which an actor is performing

An action label is associated using Fuzzy Logic to the state of the agent, depending on the agent velocity. Thus, we can distinguish between three different actions, namely *running*, *walking* or *standing*. These fuzzy attributes allow the posture status to incorporate a conceptual term, i.e. the label of the recognized action:

$$is\_performing(Agent, aLabel).$$

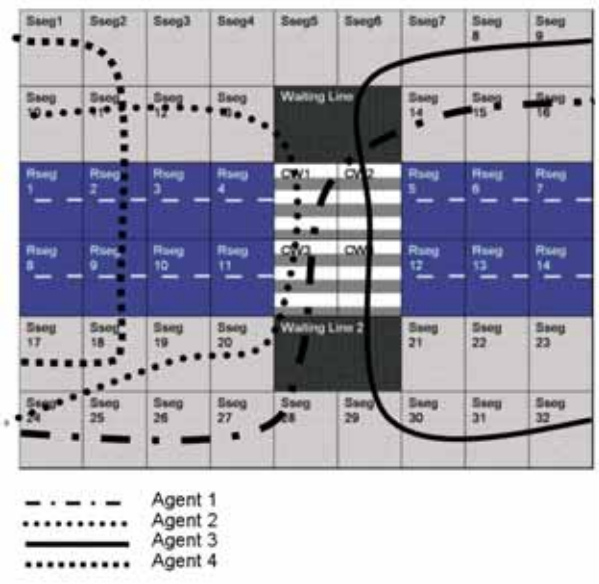

Figure 1: The roadway scene model and the agents' trajectories.

4 Analysis of Pedestrian behaviors in a roadway scene

In this section we show an example of SGT which can be used to describe human behaviour within a roadway scene. Behaviour analysis requires of an explicit reference to a spatial context, i.e., a conceptual model of the scene. Such a model allows to infer the relationship of the agent with respect to (predefined) static objects of the scene, and to associate *facts* to specific locations within the scene. All this information is expressed as a set of logical predicates in FMTL, using the language F-Limette.

The conceptual scene model used is shown in Fig. 1. The scene is divided into poligonally bounded *roadway_segments*, which describe the possible positions in which an agent can be found. Each *roadway_segment*, has a label which determines the conceptual description associated to such a segment. At present, we distinguish (at least) four different types of segments, namely: *sideway_segment*, *waiting_line*, *roadway_segment* and *crosswalk*.

Consequently, we can build predicates which relate the spatial position of the agent with respect to these segments. Our experiments are based in a recorded video sequence which involves four agents, whose trajectories are depicted in Fig. 1.

The agents' behaviors are described as:

- *Agent 1* walks through the sideway towards the waiting line and crosses the pedestrian crossing without stopping to see whether a car is approximating.
- *Agent 2* and *Agent 3* behave like *Agent 1*, but they stop in the waiting line for few seconds before crossing the crosswalk.
- *Agent 4* crosses the road without arriving to the pedestrian crossing.

Fig. 2 depicts a simplified version of an SGT which allows to infer the behaviour of agents within the roadway scene, as detailed next. The root graph comprises only one situation scheme, in which the predicate states that an agent is presently active, *active(Agent)*. The first possible specialization is the fact that the agent is not currently walking through the walking line. Then, only two situations can be instantiated: the agent is on the road or it's on the sideway. Due to in this scene there are only two kinds of segments on where an agent can appear, this situation would repeat until agent reaches the waiting line or it leaves the scene. When the agent arrives to the waiting line (*sit_ED_SIT29*) the agent might stop for checking there's no car on the road. This case is also modeled in the specialization of this situation scheme. After leaving the waiting line, agent can walk through the pedestrian crossing (*sit_ED_SIT32*) or continue walking on the sideway. Once an agent has reached the other sideway, he or she is expected to continue walking on the sideway until leaving the scene. Tables 1 and 2 show the resulting textual descriptions generated for *agent_1* and *agent_4*.

5 Conclusions and future work

We have used a deterministic model suitable for modeling human behaviors, which has been used previously in a different discourse domain, i.e. for vehicle behaviour modelling. Our information source has been an image sequence previously processed with pattern recognition algorithms, thus extracting quantitative data of the trajectories performed by human agents within a scene. Reasoning about human behavior has been achieved by means of a rule-based inference engine called F-Limette and

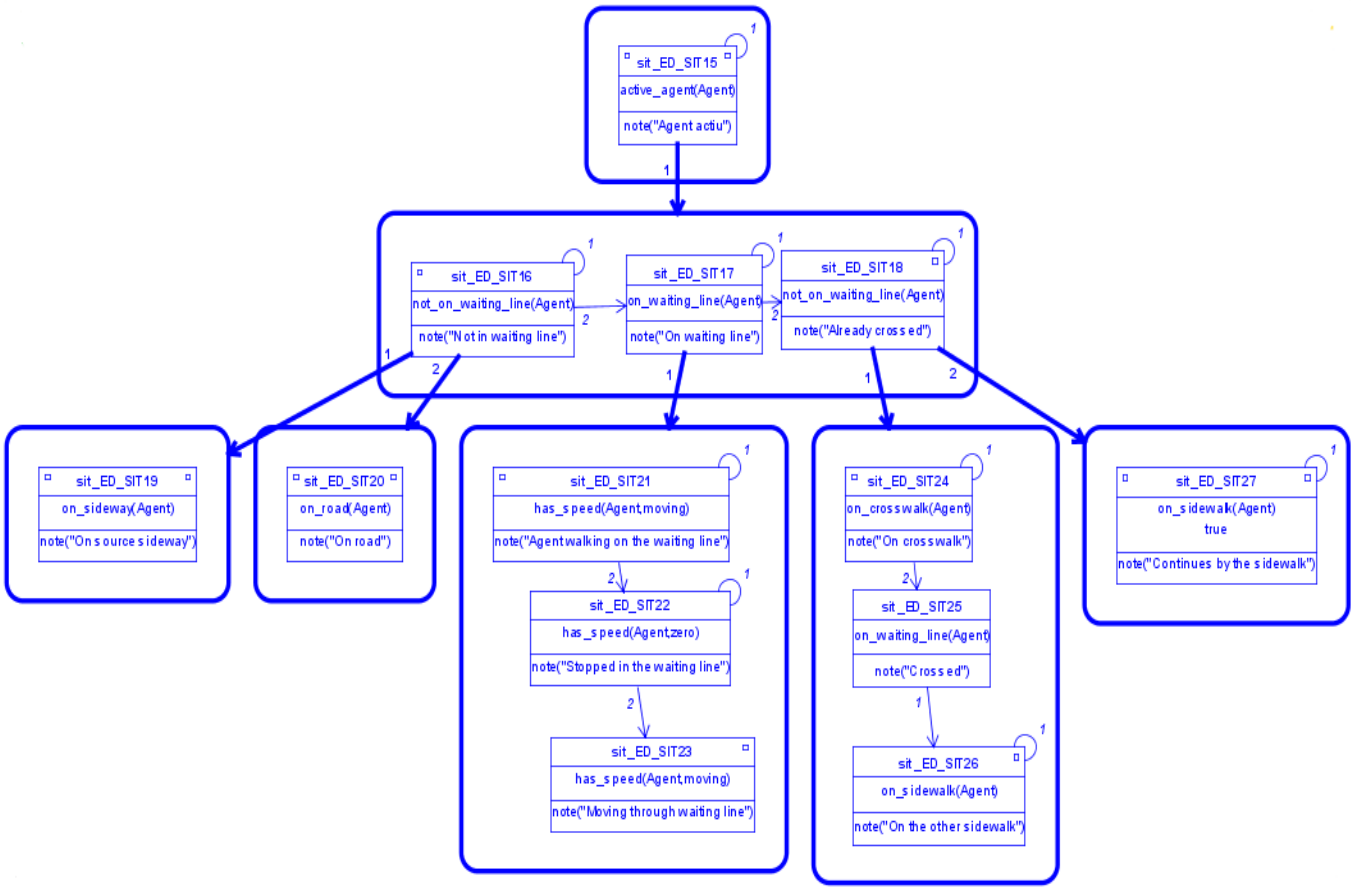

Figure 2: Situation Graph Tree for behavior interpretation of human agents crossing a roadway.

Start	End	Situation
1	26	on_sideway_seg_(**agent_1**,sseg_24).
27	76	on_sideway_seg(**agent_1**,sseg_25).
77	126	on_sideway_seg(**agent_1**,sseg_26).
127	179	on_sideway_seg(**agent_1**,sseg_27).
180	184	on_sideway_seg(**agent_1**,sseg_28).
185	225	agent_walking_on_the_waiting_line(**agent_1**).
226	321	on_crosswalk(**agent_1**).
322	371	crossed(**agent_1**).
372	521	on_the_other_sidewalk(**agent_1**).

Table 1: Sequence of textual descriptions generated for agent **agent_1** using the SGT of Fig. 2 and the trajectory data of Fig. 1.

Start	End	Situation
523	571	on_sideway_seg_(**agent_4**,sseg_17).
572	595	on_sideway_seg_line(**agent_4**,sseg_18).
596	635	on_road_(**agent_4**,rseg_9).
636	680	on_road_(**agent_4**,rseg_2).
681	740	on_the_other_sidewalk(**agent_4**).

Table 2: Sequence of textual descriptions generated for agent **agent_4**.

the Situation Graph Tree formalism as the human behavior modelling tool. This model has been tested in a street scene where the agents of interest were pedestrians. Textual descriptions have been generated which qualitatively described the observed behavior.

Future work in this materia is summarized next. At present, the SGT described here has not learning capabilities, so the accuracy of the modelled be-

havior will depend on the accuracy of the a-priory knowledge. Also, uncertainty should be handled using fuzzy predicates: this will give robustness to the system and will make it able to deal with a wider set of situations in a variety of discourse domains. We also need to provide machine learning capabilites to improve reasoning through the sets of training examples. In addition, we will increse the complexity of the behaviors expected to be interpretated by our system.

Acknowledgements

This work has been supported by the EC grant IST-027110 for the HERMES project and by the Spanish MEC under projects TIC2003-08865 and DPI-2004-5414. J. Gonzàlez acknowledges the support of a Juan de la Cierva postdoctoral fellowship from the Spanish MEC.

References

[1] M. Arens and H.-H. Nagel. Behavioral knowledge representation for the understanding and creation of video sequences. In *Proceedings of the 26th German Conference on Artificial Intelligence (KI-2003)*, pages 149–163. LNAI, Springler-Verlag: Berlin, Heidelberg, New York/NY, September 2003.

[2] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 994–999, San Juan, Puerto Rico, 1997.

[3] H. Buxton. Learning and understanding dynamic scene activity: A review. *Image and Vision Computing*, 21(1):125–136, 2002.

[4] A. Galata, N. Johnson, and D. Hogg. Learning variable-length markov models of behavior. *Computer Vision and Image Understanding*, 81(3):398–413, 2001.

[5] J. Gonzàlez. *Human Sequence Evaluation: The Key-frame Approach*. PhD thesis, Universitat Autònoma de Barcelona, Spain, 2004.

[6] M. Haag and H.-H. Nagel. Combination of edge element and optical flow estimates for 3d-model-based vehicle tracking in traffic image sequences. *International Journal of Computer Vision*, 35(3):295–319, 1999.

[7] M. Haag and H.-H. Nagel. Incremental recognition of traffic situations from video image sequences. *Image and Vision Computing*, 18(2):137–153, 2000.

[8] T. Horprasert, D. Harwood, and L. Davis. A Robust Background Subtraction and Shadow Detection. In *4th ACCV, Taipei, Taiwan*, volume 1, pages 983–988, 2000.

[9] S.S. Intille and A.F. Bobick. Recognized planned, multiperson action. *International Journal of Computer Vision*, 81(3):414–445, 2001.

[10] A. Kojima, T. Tamura, and K. Fukunaga. Natural language description of human activities from video images based on concept hierarchy of actions. *International Journal of Computer Vision*, 50(2):171–184, 2002.

[11] R.J. Morris and D.C. Hogg. Statistical models of object interaction. *International Journal of Computer Vision*, 37(2):209–215, 2000.

[12] H.-H. Nagel. From image sequences towards conceptual descriptions. *Image and Vision Computing*, 6(2):59–74, 1988.

[13] H.-H. Nagel. Steps toward a cognitive vision system. *AI Magazine*, 25(2):31–50, 2004.

[14] P. Remagnino, T. Tan, and K. Baker. Agent oriented annotation in model based visual surveillance. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, pages 857–862, Mumbai, India, 1998.

[15] D. Rowe, I. Rius, J. Gonzlez, and J.J. Villanueva. Improving tracking by handling occlusions. In *3rd ICAPR, Bath, UK*, volume 2, pages 384–393. Springer LNCS 3687, 2005.

[16] K. Schäfer. Fuzzy spatio-temporal logic programming. In C. Brzoska, editor, *Proceedings of 7th Workshop in Temporal and Non-Classical Logics – IJCAI'97*, pages 23–28, Nagoya, Japan, 1997.