Automatic Behavior Understanding in Crisis Response Control Rooms

Joris Ijsselmuiden[1], Ann-Kristin Grosselfinger[1], David Münch[1], Michael Arens[1], and Rainer Stiefelhagen[1,2]

[1] Fraunhofer IOSB, Karlsruhe, Germany
{joris.ijsselmuiden,ann-kristin.grosselfinger,
david.muench,michael.arens}@iosb.fraunhofer.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
rainer.stiefelhagen@kit.edu

Abstract. This paper addresses the problem of automatic behavior understanding in smart environments. Automatic behavior understanding is defined as the generation of semantic event descriptions from machine perception. Outputs from available perception modalities can be fused into a world model with a single spatiotemporal reference frame. The fused world model can then be used as input by a reasoning engine that generates semantic event descriptions. We use a newly developed annotation tool to generate hypothetical machine perception outputs instead. The applied reasoning engine is based on fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs), promising and universally applicable tools for automatic behavior understanding. The presented case study is automatic behavior report generation for staff training purposes in crisis response control rooms. Various group formations and interaction patterns are deduced from person tracks, object information, and information about gestures, body pose, and speech activity.

Keywords: automatic behavior understanding, smart environments, rule-based expert systems, fuzzy metric temporal logic, situation graph trees.

1 Introduction

In recent years, there has been great progress in computer vision and other areas of machine perception, for example in person tracking and body pose estimation. However, high-level systems using multiple machine perception modalities and combining multiple objects have not progressed at the same pace. We are developing a toolkit for automatic behavior understanding that deploys multimodal machine perception for multiple objects, fuses everything into a world model with a single spatiotemporal reference frame, and generates semantic descriptions about the observed scene. The current system uses a dedicated annotation tool instead of multimodal machine perception as shown in Figure 1.

The presented case study is situated at the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen, during one of their staff exercises

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 97–112, 2012.

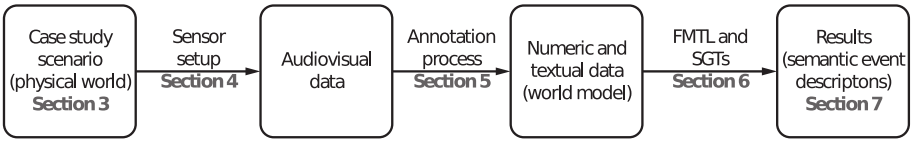

Fig. 1. System overview

for crisis response control room operations (see Figure 2). The task is to automatically generate behavior reports from multimodal machine perception during staff exercises and actual crisis management. These reports about staff behavior in the control room can be used for training purposes, evaluations, and audit trails. For instance, given the identity, position, orientation, and speech activity of the staff members over time, and information about objects in the room, these reports can contain descriptions and visualizations of group formations and interaction patterns, i.e. who was doing what with whom, using which support tools. This can be combined with audiovisual recordings and visualizations, and with the corresponding developments in cyberspace, i.e. field unit status, crisis dynamics, and other context information. Such a system would provide a rich information source, conveniently searchable for specific events.

The presented reasoning process is domain independent because it is separated from any machine perception it might use. The annotation process too is designed to be customizable for other application domains. Possible application domains include other behavior understanding applications, multimedia retrieval, robotics, ambient assisted living, intelligent work environments, intelligent user interfaces, indoor and outdoor surveillance, and situational awareness and decision support for military and civil security. Applied machine perception can range from video to radar, and from person and vessel tracking to body pose estimation, speech recognition, and activities in cyberspace. Other uses include camera control, sensor deployment planning, future event prediction, information exchange between system components, and top-down knowledge for machine perception to guide its search and improve outputs.

This paper is organised as follows. After discussing related work in Section 2, we explain the applied processing chain step by step as depicted in Figure 1. Section 3 describes the case study scenario: automatic behavior report generation for training purposes in crisis response control rooms. A staff exercise was recorded using multiple cameras and microphones with appropriate postprocessing as described in Section 4. The next step is turning the recorded audiovisual data into a world model consisting of numeric and textual data. Ultimately, this should be accomplished using machine perception and multimodal fusion, but we currently use a different approach. Section 5 describes how the postprocessed audiovisual data was manually analysed and annotated using a tool specifically developed for such purposes. The resulting world model forms the input for the reasoning engine based on fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs) presented in Section 6. It delivers semantic descriptions about staff behavior, which can be compared to ground-truth results annotated