Prometheus: Prediction and interpretation of human behaviour based on probabilistic structures and heterogeneous sensors

Jörgen Ahlberg[1], **Dejan Arsić**[2], **Todor Ganchev**[3], **Anna Linderhed**[4], **Paulo Menezes**[5], **Stavros Ntalampiras**[6], **Tadeusz Olma**[7], **Ilyas Potamitis**[8], **Julien Ros**[9]

1 Scope

The on-going EU funded project Prometheus (FP7-214901) aims at establishing a general framework which links fundamental sensing tasks to automated cognition processes enabling interpretation and short-term prediction of individual and collective human behaviours in unrestricted environments as well as complex human interactions. To achieve the aforementioned goals, the Prometheus consortium works on the following core scientific and technological objectives:

1. sensor modeling and information fusion from multiple, heterogeneous perceptual modalities;
2. modeling, localization, and tracking of multiple people;
3. modeling, recognition, and short-term prediction of continuous complex human behavior.

2 Technology

The Prometheus technology is based on the use of a network of heterogeneous sensors, the data streams of which are fed to fully probabilistic framework. This framework performs adaptive fusion of heterogeneous sources of information, which involves integrating information - in the broadest sense - to detect, estimate and predict the global state of interacting people (please refer to Fig.1). The redundancy and complementarity of information provided by the heterogeneous sensors facilitate the robust estimation of states and interpretation of behaviours. Moreover, in Prometheus project the sensor set was chosen in an manner which enables surpassing the weaknesses of each perceptual modality in dealing with coverage of the sensed area and its response to occlusion, noise and differing environmental conditions. The description of the multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data from multiple perceptual modalities comprises of data pre-processing and feature extraction followed by a hierarchy of four processing levels. These higher processing levels are: (i) Object Assessment, (ii) Situation Assessment, (iii) Impact Assessment and (iv) Sensor Management. The logical flow among the processing levels is presented in Fig.2.

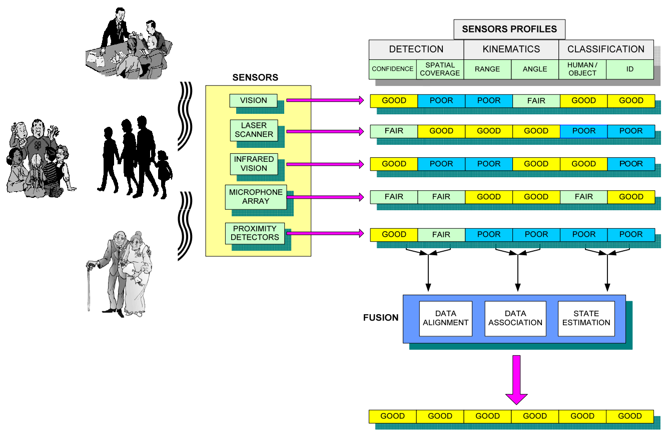

Figure 1. Fusion of heterogeneous sensors, which offer complementary information, allows for robust estimation

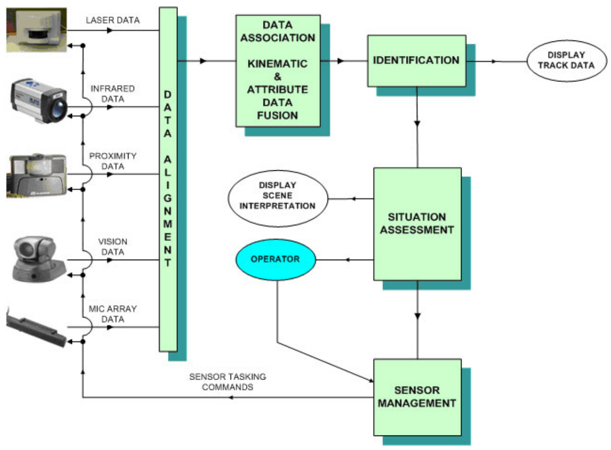

Figure 2. Overall system architecture

[1] Swedish Defence Research Agency (FOI), Sweden, email: jorahl@foi.se
[2] Munich University of Technology (TUM), Germany, email: arsic@tum.de
[3] University of Patras (UOP), Greece, email: tganchev@wcl.ee.upatras.gr
[4] Swedish Defence Research Agency (FOI), Sweden, email: anna.linderhed@foi.se
[5] University of Coimbra (FCTUC), Portugal, email: paulo@isr.uc.pt
[6] University of Patras (UOP), Greece, email: dallas@wcl.ee.upatras.gr
[7] MARAC S.A., Greece, email: olma@marac.gr
[8] Technological Educational Institute of Crete, Greece, email: potamitis@stef.teicrete.gr
[9] Probayes SAS, France, email: julien.ros@probayes.com

From: Proc. ECAI 2008, Patras, Greece. ECCAI

3 Applications

While the proposed framework is of general purpose, the technology developed in Prometheus project will be implemented and demonstrated in a number of scenarios related to the surveillance of large shared spaces or human-friendly human-machine-interaction in home environment. Depending on the specific problem at hand, a variety of applications can take advantage of the Prometheus technology. Typical examples, can be in the fields of security (detection of abnormal behaviour), home/health care for elderly people, video analysis of sport activities; point-of-sales advertising, or customer audience analysis, etc.

4 Demonstration

An important part in the Prometheus project aims at building up a large database of multi-sensor data that can be used for development of new algorithms and creation of models of the physical environment. Besides typical human-to-human and human-to-object interactions the database contains examples of abnormal behaviours, such as thefts in action, street fights, etc. Complex storyboards and taskcards are created, so that actors get precise orders where to go, which route to take and what kind of activity to show. This way of planning facilitates the database annotation.

One of the first problems encountered in a surveillance application using an heterogeneous network of sensors is to detect [4] and track [2] several persons under severe occlusion and clutter. To address these difficulties, the Prometheus database is recorded in both outdoor and indoor environments, utilizing a heterogeneous set of sensors: (i) High Resolution Cameras for scene overview and detail views, (ii) 3D cameras, (iii) Thermal Infrared Camera, (iv) Microphone Arrays. Especially the outdoor scenario was selected to provide a close match to real-life situations, with changing lighting conditions, moving background, shadows and a varying number of people within the scene, interferences form the environment, etc. An additional challenge comes from the high environmental temperatures, which interfere with the resolution of the thermal sensors. Fig. 3 illustrates two views of the outdoor data recording site.

Figure 3. Example view of the outdoors data recording site

These realistic data will be used to demonstrate state-of-the-art algorithms and improvements of already existing algorithms. For instance, work presented at the PETS 2007 workshop [1] is illustrated in Fig. 4). Consequently, we present the first results about how our pedestrian tracking technology operates on real-world data, such as those recorded during the initial recording campaign.

A second point is to handle the case where a person re-enters the scene after a while and so in the field of view again. The task is finding an efficient algorithm, which is capable to add a new individual to an existing database, update the model for different poses and recognizes people reappearing in the field of view. Therefore scale invariant features have to be extracted from the person's body and stored in a model. First trials with so called SIFT features showed promising results even from different field of views. In an ideal implementation it will be tried to create an 3D model from the person either from one or more fields of view. A sample for matching points on a person's body is displayed in Fig. 5.

Figure 4. Example from the 2007 PETS workshop

Furthermore, an automatic analysis of typical scenes is performed using probabilistic approaches, such as shown in [3]. Results with a heuristic approach for left luggage detection are shown in Fig. 4.

In addition, preliminary results of a behavior modeling technique will be demonstrated for specific scenarios. This includes the demonstration of combined tracking and interpretation techniques. To illustrate this, during the recording campaign we recorded complex scenarios of people interacting with other people or objects such as ATM machines.

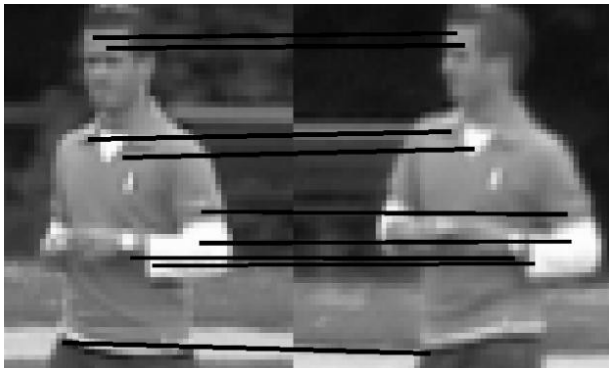

Figure 5. Matching of feature points

REFERENCES

[1] D. Arsić, M. Hofmann, B. Schuller, and G. Rigoll, 'Multi-camera person tracking and left luggage detection applying homographic transformation', in *Proceeding Tenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2007, IEEE, Rio de Janeiro, Brazil*, (October 2007).

[2] S. Calderara, R. Cucchiara, and A. Prati, 'Bayesian-competitive consistent labeling for people surveillance', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**(2), 354–360, (2008).

[3] N. Carter, D. Young, and J. Ferryman, 'A combined bayesian markovian approach for behaviour recognition', in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pp. 761–764, Washington, DC, USA, (2006). IEEE Computer Society.

[4] C. Papageorgiou and T. Poggio, 'A trainable system for object detection', *Int. J. Comput. Vision*, **38**(1), 15–33, (2000).