

Sensor Fusion and Environmental Modelling for Multimodal Sentient Computing

Christopher Town
University of Cambridge Computer Laboratory
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
cpt23@cam.ac.uk

1 Overview

Sentient computing uses networks of sensors to capture and maintain an internal representation (“world model”) of an indoor environment, thereby allowing applications to have greater awareness of users and their requirements. This chapter shows how computer vision information obtained by several cameras can be used to enhance the capabilities of a sentient computing system which previously relied on ultrasound to track people and devices. Integration is achieved at the system level through the metaphor of shared perceptions in the sense that the different modalities are guided by and provide updates for a shared internal model. This world model incorporates aspects of both the static (e.g. positions of office walls and doors) and dynamic (e.g. location and appearance of devices and people) environments. It serves both as an ontology of prior information and as a source of context which is shared between applications. Fusion and inference are performed by Bayesian networks which model the probabilistic dependencies and reliabilities of different sources of information over time. It is shown that the fusion process significantly enhances the capabilities and robustness of the system, thus enabling it to maintain a richer and more accurate world model. Further details are available in [2, 3].

2 The Sentient Office

A sentient office uses sensor and resource status data to maintain a model of the world which is shared between users and applications. Sensors and telemetry are used to keep the model accurate and up to date, while applications see the world via the model.

The SPIRIT [1] system uses mobile ultrasonic sensor devices known as “Bats”. The achieved spatial granularity is better than 3cm for $> 95\%$ of Bat observations (assuming only small motion) and Bats may be polled using radio base stations and a variable quality of service to give update frequencies of up to 25Hz while remaining scalable to hundreds of tagged people and devices in a large office. The

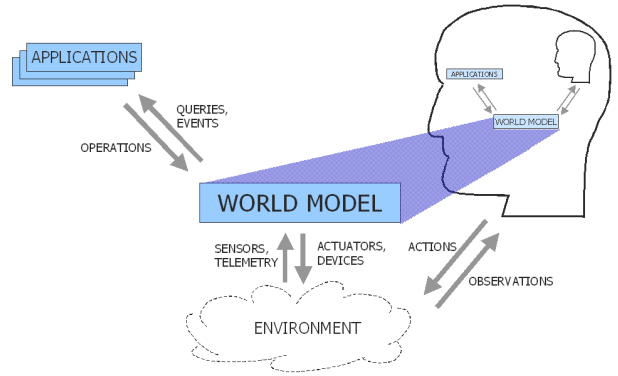


Figure 1: **Diagrammatic overview of the world model maintained by the sentient computing system.**

system maintains an internal dynamic representation of the office environment and objects and interactions which occur within it. This world model (see 1) comprises a static part consisting of those aspects of the environment which are not monitored using the ultrasonic location system and a dynamic part consisting of those objects and devices which are. A richer and more accurate model enables applications to better perceive context and thereby interact with users in a more natural way.

3 Adaptive Multi-modal Fusion of Tracking Hypotheses

The dynamic component of the world model benefits from a high-level fusion of the visual and ultrasonic modalities for robust multi-object tracking and appearance modelling. The approach taken here uses a Bayesian graphical network (shown in figure 3) to integrate information from the different sources. Discrete reliability indicator variables (R_S , R_F , R_D , R_C , and R_B) are used to model how reliable each modality is at the current time. The network serves as a shared template from which individual tracking hypotheses are derived. Hypotheses are instantiated by SPIRIT ob-

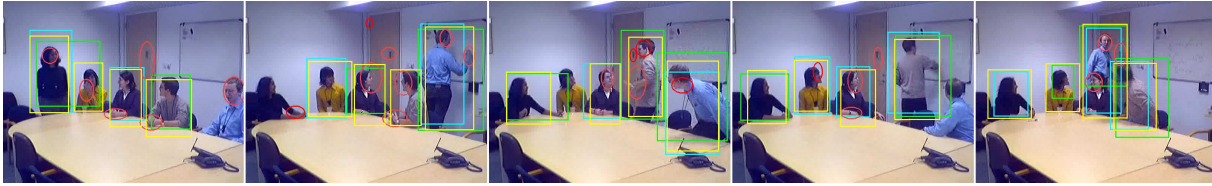


Figure 2: **Tracking results: Rectangles denote object hypotheses derived from the two modalities (green: visual tracking, blue: SPIRIT observations) and the results of fusion (yellow). Red ellipses indicate face detections.**

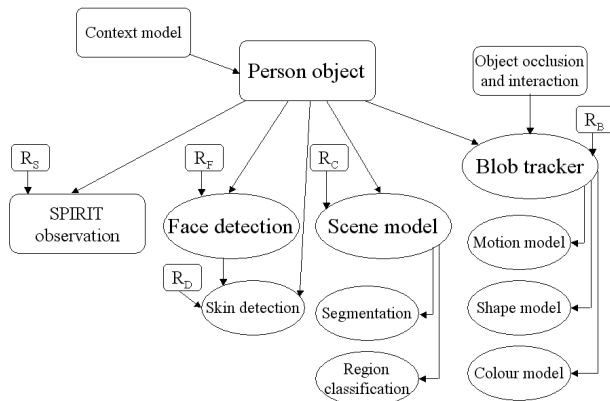


Figure 3: **Bayesian graphical model for multi-modal fusion. Reliability variables allow adaptive integration of different sources of information.**

servations or a visual blob tracking framework, thus allowing tracking of people who are not tagged with a functioning Bat device or who are not currently visible by a given camera. Other visual cues such as skin colour and face detection function as supporting modalities. Spatial and object-specific ontologies from the world model or the region segmentation and classification methods provide contextual constraints and guide the generation of hypotheses. Reliabilities are adapted on the basis of manually specified rules over reliability indicators such as motion and appearance variation and performance feedback measures such as consistency and log-likelihood of the observations under each modality. Posterior probabilities for each hypothesis can then be computed by integrating all available information using the fusion network.

Each hypothesis maintains its own set of reliability variables and competes for observations with other hypotheses. The conditional probabilities (including the dependency on reliability variables) of the underlying network structure were estimated by means of the EM algorithm from manually labelled training sequences. Temporal evolution of the model occurs via a Kalman or particle filter applied to a colour blob tracker and through the modification of reliability variables in light of current observations. Some results from a meeting room sequence are shown in figure 2.

4 Environmental Modelling

In order to improve the world model's representation of personnel, the following tasks are being addressed by fusing SPIRIT and computer vision information (see [2, 3] for further details):

- 1.) *Enhanced tracking*: making position estimates more robust and accurate
- 2.) *State estimation*: modelling appearance and inferring orientation and body posture
- 3.) *Identification*: detecting potential intruders (people not wearing functioning Bats) and providing vision-based biometric verification of identity

The world model's static component can also be augmented using vision-based environmental modelling techniques. In particular, it is possible to perform:

- 1.) *Enhanced visualisation*: the visualisation of rooms can be made more compelling by acquiring texture maps of actual features such as decorations, carpet colouration, and white board surfaces
- 2.) *State estimation*: aspects of environmental state such as that of doors (open, closed, ajar) and lights (on, off) can be determined
- 3.) *Detection*: semi-static parts of the environment such as furniture and computer screens can be recognised using vision techniques

References

- [1] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggle, A. Ward, and A. Hopper. Implementing a sentient computing system. *IEEE Computer*, 34(8):50–56, 2001.
- [2] C.P. Town. *Ontology based Visual Information Processing*. PhD thesis, University of Cambridge, 2004.
- [3] C.P. Town. Multi-sensory and multi-modal fusion for sentient computing. *Int. Journal of Computer Vision*, 71(2):235–253, 2007.