In this paper a new method of detecting and tracking a human person in three dimensional space using audio and video data is proposed. A simple tracking system with two microphones and stereo vision is introduced. The audio information is resulting from the generalized cross correlation (GCC) algorithm, and the video information is extracted by the continuously adaptive mean shift (CAMshift) method. The localization estimates delivered by these two systems are then combined using a novel particle swarm optimization (PSO) fusion technique. In our approach the particles move in the 3D space and iteratively evaluate their current position with regard to the localization estimates of the audio and video module. This facilitates the direct determination of the object’s three dimensional position. Compared to existing methods, this novel technique achieves faster tracking performance while being independent of any kind of model, statistic, or assumption.

Stichworte:
- audio-visual systems
- feature extraction
- image fusion
- object detection
- particle swarm optimisation
- stereo image processing
- audiovisual fusion
- continuously adaptive mean shift method
- generalized cross correlation algorithm
- human person detection
- localization estimates
- particle swarm optimization
- stereo vision
- three dimensional object tracking

Kongress- / Buchtitel:
- Information Fusion, 2008 11th International Conference on

Jahr: