Real-Time Fusion of Multimodal Tracking Data

Martin Weser (diploma thesis)

email: weser@informatik.uni-hamburg.de

Introduction

Service robots will increasingly support daily work in business or home environments in the near future. Possible services are delivery tasks, cleaning services or home care. However, the distribution and thus further development of mobile robots is mainly dependent on the acceptance of society. An important criteria for this acceptance is the robot's ability to interact with the environment. Therefore it is essential to give the robot a detailed model of its environment, i. e. the location of its interaction partners. In general, this knowledge can only be generated using sensory input. An explicit specification of a dynamic environment is usually impossible.

Multimodal Tracking Framework

Both camera tracking as well as laser tracking have their own specific advantages and drawbacks. To build a robust and accurate tracking system it is necessary to integrate independent tracking algorithms working on different sensor modalities. With an appropriate fusion algorithm the specific advantages of the sensors could complement one another to decrease the overall error.

A typical multiple target tracking system consists of four blocks: sensor hardware, single sensor tracking, data fusion and association and track life management. A tracking system should be modular to allow addition, removal and exchange of sensors and tracking algorithms. Therefore, the most important aspect of a tracking system is its ability to filter and fuse the results from individual sensors.

I developed a framework that contains the above-mentioned blocks. The schematic block diagram for this framework is shown in fig. 1. The implementation of the data fusion block includes filtering and data association using a particle filter. Through this interface an arbitrary number of tracks is provided where each track consists of (a) the current position and velocity of the tracked object, (b) the uncertainty about the position and velocity and (c) a unique identity number.

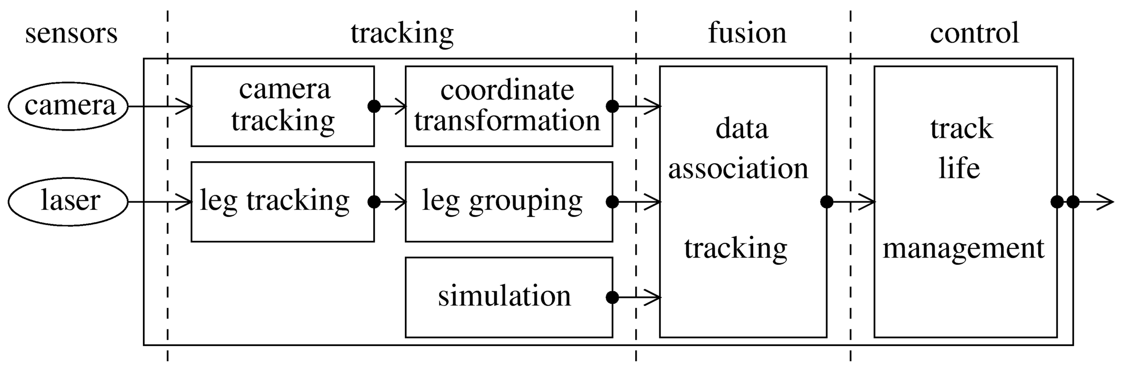

Figure 1: Components of the tracking system. The black dots denote the common interface.

Sensor Fusion and Filtering

The problem of tracking can be considered as the detection of the state of a target. Therefore, the state x_t of a tracked person at time t is modeled as a four-dimensional vector $(x, y, \delta x, \delta y)^T$. This vector not only describes the position on the ground plain but also the velocity of the person. Since measurements of sensors contain errors it is impossible to derive the actual state of observed persons in a non-probabilistic way. Generally, a probability density function (*pdf*) is used to represent the state. Nonlinear Bayesian filtering can be applied to determine this *pdf* by taking every previous measurement into account, however practically it can only be applied when certain constraints hold. The Kalman filter and the particle filter are two frequently used methods to realize bayesian like filtering.

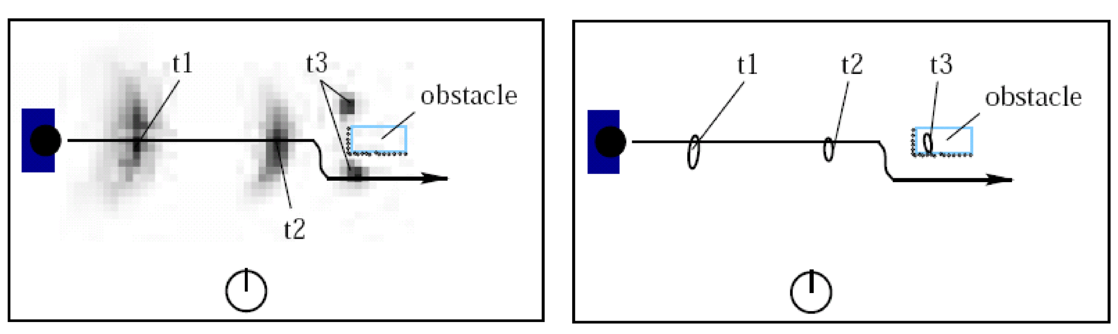

Figure 2: A person is trackted by a particle filter (left) and a Kalman filter (right). [Schulz *et al.* IJRR 2003]

The Kalman filter is not able to handle the nonlinear *pdf* shown in fig. 2 thus the particle filter is used here.

The basic principle of the particle filter is the *importance sampling*. A multidimensional function $g(x)$ is factorized into two functions $g(x) = f(x)\pi(x)$, where $\pi(x)$ is interpreted as a probability density function with $\pi(x) \geq 0$ and $\int \pi(x)\,dx = 1$.

If a set of samples $\{x^i | i = 1, \ldots, i = N\}$ with $N \gg 1$ and distributed according to $\pi(x)$ is generated, the integral of the function $g(x)$ can numerically be approximated as

$$\int g(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} f(x^i) \qquad (1)$$

Here the function $g(x)$ is the state of the tracked person.

Camera-Based Tracking

In this system the approach presented by Comaniciu (2000) is used since it is suitable for cameras mounted on a mobile robot. People tracked in the camera image are represented by a weighted color histogram. Pixels are weighted with a monotone decreasing kernel function $K : \mathbb{R}^2 \to \mathbb{R}$ which assigns smaller weights to the pixels which are further from the center of a detected person. If the size of a person is denoted by $2h^*$, the probability of the object's color u can be calculated as follows:

$$\hat{q}_u = C \sum_{x \in X^*} K\left(\frac{x}{h^*}\right) \delta\big(b(x) - u\big), \qquad (2)$$

where C denotes a normalization constant. The function $b(x)$ assigns the pixel to an index of the histogram's color bin. Therefore, a person located at the coordinate y in the image plane is represented by a color histogram:

$$\hat{p}_u(y) = C_h \sum_{x \in X_h^y} K\left(\frac{y - x}{h}\right) \delta\big(b(x) - u\big) \qquad (3)$$

where h is the size of the target candidate. As a measure of similarity between two color histograms we chose the *Bhattacharyya coefficient*.

The goal of the tracking algorithm is to find the location y with the highest similarity between the color histogram of a person and a candidate located at y.

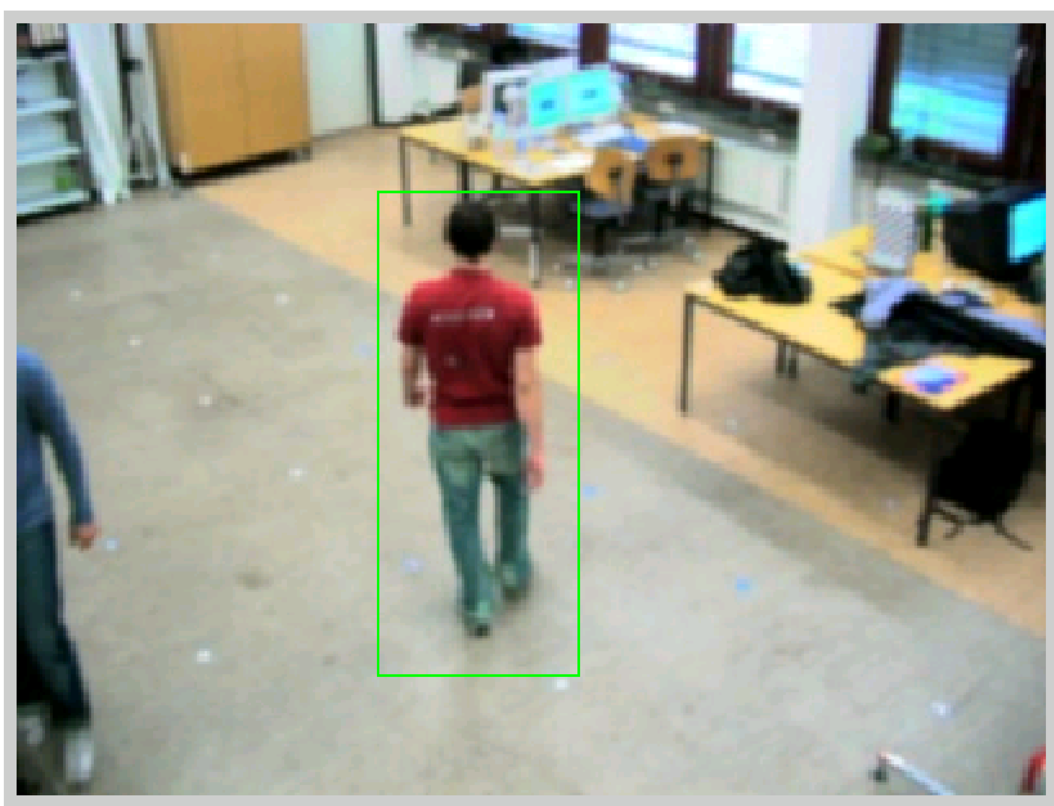

Figure 3: A person is tracked in an image.

Laser-Based Tracking

Tracking algorithms which use laser range finders are often divided into two steps. Here a novel method to estimate the background distances is used which is updated with each measurement. The background distance $h_i(t)$ at time t and angle i is given by the following recursive equation:

$$h_i(t) = h_i(t-1) + \begin{cases} \epsilon_1 & \text{if } h_i(t-1) < m_i(t) \\ (-\epsilon_2) & \text{else} \end{cases}, \qquad (4)$$

where $m_i(t)$ is the measurement at time t. The values of the positive increments ϵ_1 and ϵ_2 determine the adaptivity of the background model. Background measurements are removed afterwards and groups of foreground measurements are tracked with a particle filter.

Experimental Results

For experiments two SICK laser range finders mounted on a mobile service robot and a stationary camera at the laboratory of the TAMS institute are used. Due to the uncertainty of the camera tracking which is caused by noisy measurements and changing illumination conditions the outcome of the laser tracking is weighted higher. In figure 4 the observed person's true trajectory is assumed as linear. Although the greater variance of the trajectory computed by the camera algorithm is obvious, the fused result has be improved compared to the laser tracking result.

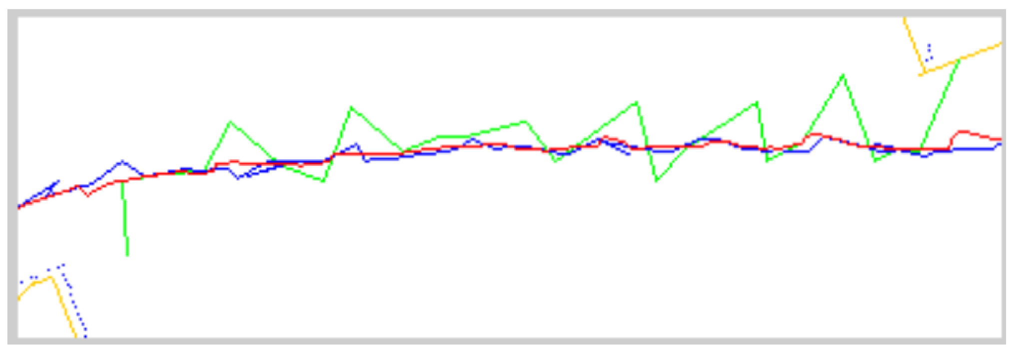

Figure 4: Comparison of sensor modalities: Camera tracking (green) and laser tracking (blue) are fused by a particle filter (red).

Both sensor modalities are used to increase the accuracy and robustness of the tracking algorithm. Figure 5 shows a multimodal tracking of two persons. Figure 6 shows the results of a two-hour observation of the TAMS floor.

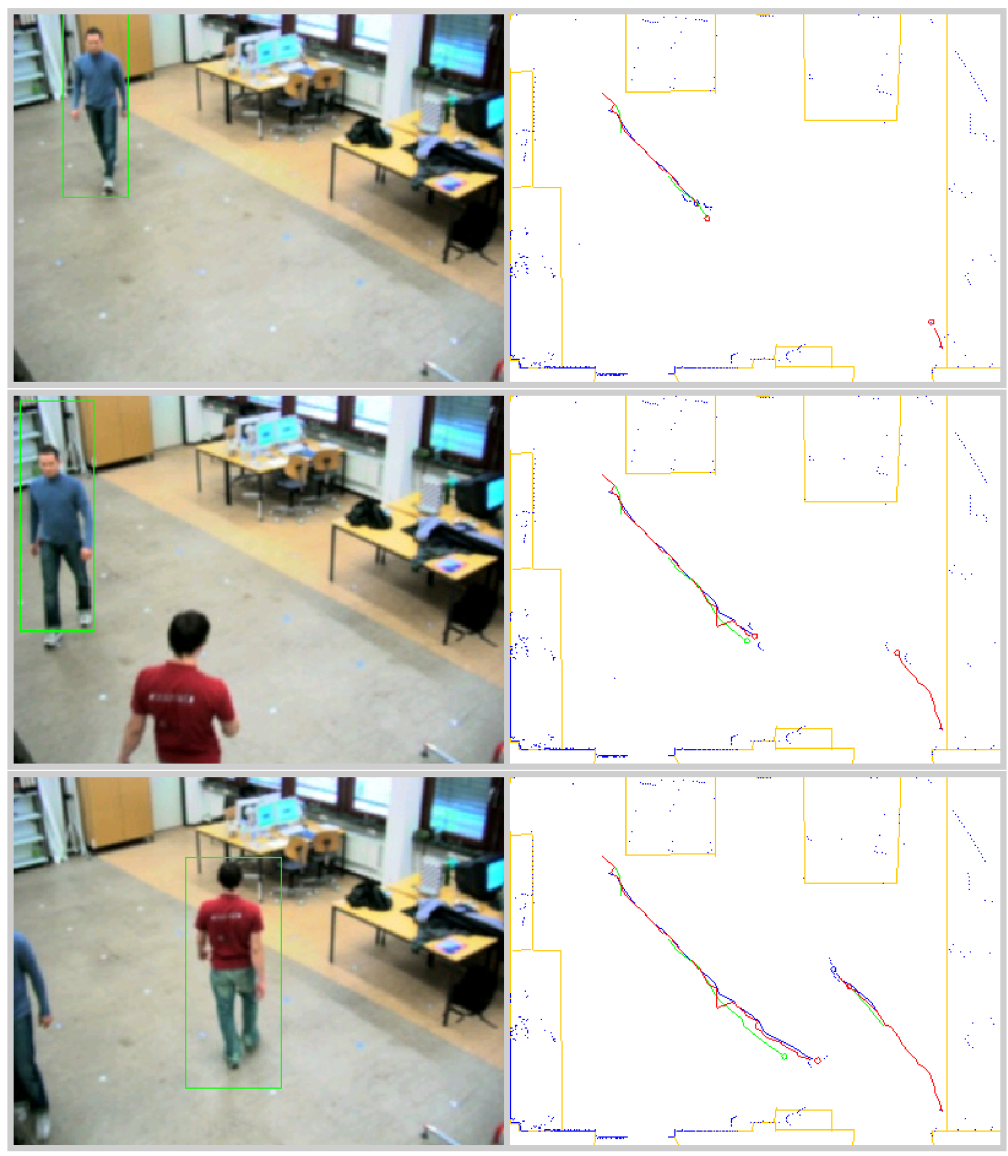

Figure 5: Increased robustness due to the use of multimodal sensors.

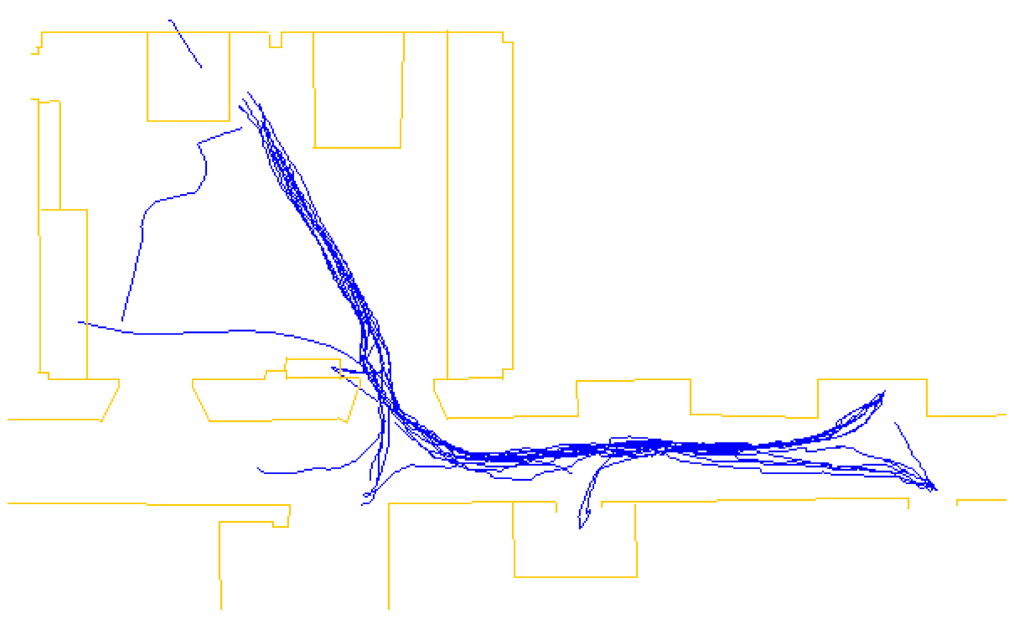

Figure 6: Results of a two-hours observation. The tracking starts when persons become visible to the sensors and ends when they leave the range of the sensors.

The camera-based tracking algorithm runs with a resolution of 640x480 pixel. With a standard pc our implementation achieves 25 fps while tracking 3-4 targets. The laser-based algorithm reaches up to 30 fps due to the lower amount of data. Since the used particle filter is an efficient approximation of the bayesian filtering, our system should be able to ensure real-time constraints if an appropriate environment, e.g. RTLinux, is used.