 In Proc. of 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Statistical Recognition of Motion Patterns

Jörg Illmann, Boris Kluge, Erwin Prassler

Research Institute for Applied Knowledge Processing
Helmholtzstr. 16, D-89081 Ulm, Germany
{illmann, kluge, prassler}@faw.uni-ulm.de

Abstract

This paper addresses the problem of people detection in dynamic environments in real-time and the extraction of typical motion patterns. The need for estimation of peoples movements with a mobile robot is essential for many applications in the field of service robotics. One key prerequisite for a machine to interact intelligently with people is its ability to recognize humans and their behavior. This paper presents an approach for tracking people in dynamic environments and to determine their behavior.

1 Introduction

Essential for a tracking system which operates in dynamic environments and interacts with people is its robustness. The system has to be robust with respect to continuous changes of the appearance of objects and the environment. Changes of lighting conditions must not affect the tracking performance. In addition the system has to be able to evaluate whether the conditions are adequate for object tracking and which method and configuration match best to the current situation. Another requirement is that the system be able to cope with temporal occlusion of the object by static or dynamic parts of the environment. For an intelligent interaction with people in the environment it is necessary to be able to track multiple objects concurrently in real-time and to determine their behavior. And of course the system should operate in real-time.

We tackled this problem using a bunch of different tracking procedures. Two of them, both relying on spatial knowledge are presented. In the first approach people are recognized by means of spatial and color information which is acquired by a multi sensor system consisting of a laser range finder and an omnidirectional color camera. Spatial knowledge is used to partition the scene into subparts which are considered in the tracking process. The color distribution of each of these subparts is analyzed and depending on the distribution an appropriate tracking procedure is chosen. To provide robustness in different scenarios color and range information is fused.

The second approach solely relies on spatial information [1]. Objects are extracted from laser range finder images applying simple heuristics to partition the scene into primitive parts. Correspondences between successive scans are established by network optimization techniques.

The estimated trajectories are used to extract basic motion patterns like straight motion, wandering around aimlessly, entering a queue or just waiting for something. A tree structured vector quantization is applied to the persons trajectories to restrict the number of elementary movements. A time series of the quantization results is fed into an un-supervised classification stage which acquires a representation of characteristic motion patterns.

The main topics in this article are fusion of omnidirectional vision with range information provided by a laser scanner, color segmentation of omnidirectional images, object tracking and prediction of object trajectories.

1.1 Related Work

Different approaches for tracking of persons can be divided into methods based on moving cameras and methods which rely on stationary cameras and stationary background.

Much work on tracking systems for stationary cameras originates from virtual reality applications where persons move in front of a stationary background. Typically these systems segment images using frame differencing [5].

These approaches are not suitable for tracking on a mobile robot since figure-ground separation can not be accomplished by simple image differencing. Wren et al. [6] represent a person by a collection of colored blobs. These blobs and the scene background are modeled by a three dimensional normal distribution.

Detection of a person can be easily achieved by color classification. While their blob representation is useful they require a smoothly changing background. This can not be guaranteed on a moving platform.

In the last years there has been an increasing interest in using range sensors for tracking of objects since the spatial arrangement of objects is often more relevant than appearance information. An approach for a driver assistance system where objects on motor ways are continously tracked is proposed by [2]. Sobottka [4] use a feature-based approach to obstacle tracking in range image sequences of traffic scenes. They strongly rely on three-dimensional sensory information to find correspondences between features in consecutive frames.

2 System Description

Our experimental platform MAid (see Fig. 1) is a robotic wheelchair which was equipped with a variety of different sensors. It is based on a commercial electrical wheelchair type SPRINT manufactured by MEYRA GmbH in Germany.

An on-board 200 MHz processor handles path planing, obstacle avoidance and motion execution in addition to the tracking mechanisms.

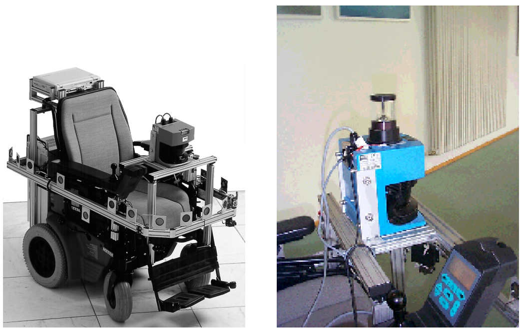


Figure 1: *Intelligent transportation system MAid*

MAid's sensor apparatus includes

- a *dead-reckoning system* consisting of a set of wheel encoders and a optical fiber gyroscope,
- a *modular sonar system* consisting of 3 segments each equipped with 8 ultra-sound transducers,
- a SICK 2D *laser range finder* LMS 200 mounted on a removable rack.
- an *omnidirectional color camera* mounted above the laser range finder.

The sonar system and the laser range finder are the main sources MAid uses to actually perceive the surrounding environment. This perception consists of two dimensional range profiles and gives a rather coarse picture of the environment. The range profiles are used to extract the basic spatial structure of the environment. In particular, MAid uses these range data to detect and to track moving objects in its surroundings. Right above the laser range finder an omnidirectional camera is mounted. To facilitate the integration of range data with vision data the central axis of the camera aligns with the rotation axis of the range finder.

3 Tracking by Fusion of Range- and Color-Information

In this section we present a new approach for tracking of multiple people in dynamic environments with a moving observer. Most approaches for tracking reported in the literature assume a static camera [5] or a known scene background. Thus moving objects can be easily detected simply by differencing current images and known background. Many tracking procedures for moving observers rely on known color features or markers on the person. A common problem is that they are susceptible to illumination changes especially in outdoor scenes.

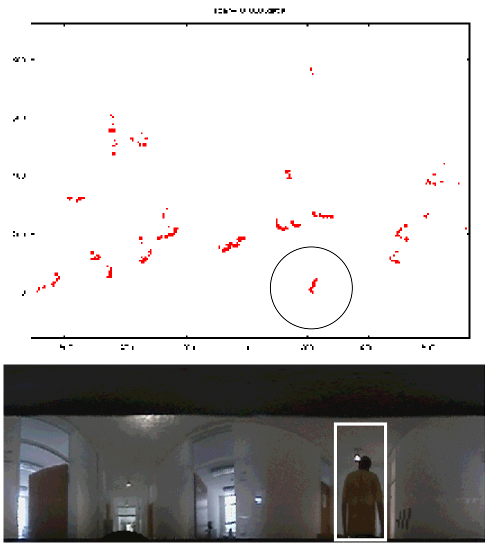
This paper concentrates on tracking of people in different environments with a moving observer. We treated this problem by using range and appearance measuring sensors. While range information can be estimated very robustly by conventional laser range finders, color vision is typically vulnerable to illumination changes. On the other hand color and visual appearance are good features to differentiate between people. Therefore we rely on robust person detection based on robust range information and use the color mainly for differentiation and validation. Finally both kinds of information are fused to generate person hypotheses.

3.1 Range Based Scene Segmentation

Figure 2 shows a typical scan of a scene generated by a laser range finder. The predicted position of the person defines a *Region of Interest* ROI, shown by a circle. Objects detected within this Region of Interest are candidates for the person to be tracked.

Once we have an expectation where the person can be found in the environment we can use this spatial information to segment the omnidirectional image. Transferring the spatial information from the laser range scan to the image is quite simple if the rotation axis of the scanner and the optical axis of the camera coincide. After an initial calibration of the omnidirectional camera angular values can be transformed into pixel positions in the image.

In this early stage of processing we are not trying to detect persons in the range scan directly (like fe. [1]).



(a) range scan (b) segmented color image

***Figure 2:** Range segmentation*

Instead we transfer every measurement lying within the Region of Interest and tag the associated pixels in the image. This pixels are considered in further image processing steps.

The whole tracking process is depicted in figured 3.

3.2 Startup phase

Whenever an object enters the observed scene an appearance model of the object is created. This model is used during the tracking phase to keep track of the object. Our object representation consists of geometric features describing the objects extent, position and motion and appearance features describing the color distribution.

New objects are detected by segmenting the scene using range information provided by the laser range finder. All objects which are currently not known are new tracking objects. The result of the segmentation in the range image is used to segment the omnidirectional color image by focusing on an image area which corresponds to the candidate region.

An object model for tracking is acquired by transformation of the color image to a Hue-Saturation-Intensity (HSI)-color space representation. This representation has proven to be quite robust to intensity changes which are caused by changing light conditions. Illumination changes can be caused when the tracked person enters a shadowed area. To ignore the intensity we use only the hue-saturation subspace.

Depending on the current light conditions and color distribution properties of the target object an appropriate tracking mechanism is chosen. If the object does not contain colors which allow to differentiate it from the background (fe. white) we use a Haussdorf-metric based person tracking [3]. If the distribution is unimodal the distribution is represented by fuzzy rules, describing hue and saturation of the person. If it contains multiple colors a histogram based scene segmentation is used.

In addition to the selection of a tracking process the light conditions are monitored. If they are to bad for visual tracking the process is stopped and a warning is output.

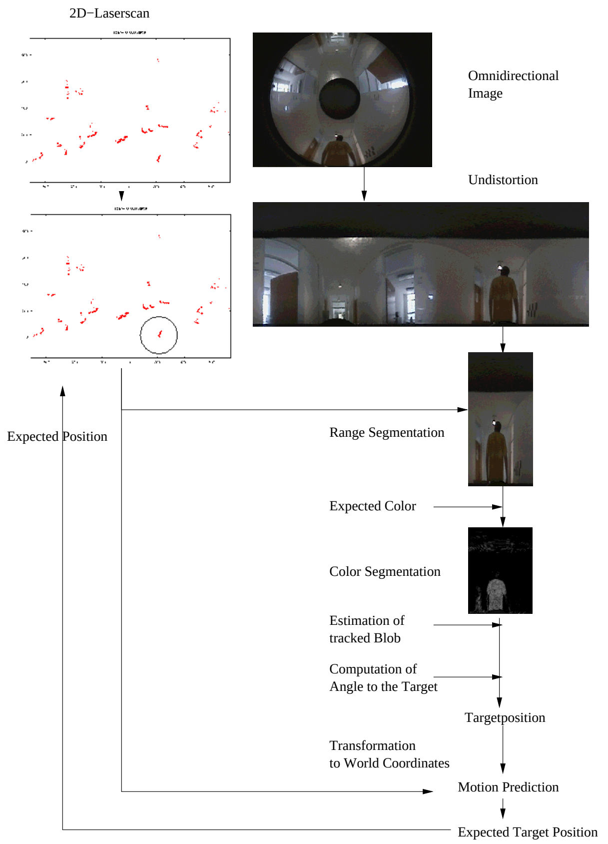


***Figure 3:** Tracking procedure*

3.3 Color Histogram Tracking

To track objects we use a statistical approach based on back-projection of an extracted color histogram. Color histograms describe the contribution of color components. This color distribution is characteristic for the color to be tracked.

All the pixels marked during the range based segmentation stage are analyzed for their color components. If the color does not have high probability to belong to the target distribution it is eliminated. A morphological step is applied to the remaining pixels, eliminating single defects. Candidate regions for the target are estimated by region-growing on this representation.

4 Range Based Tracking

This section presents an object tracking system based on range information provided by a laser range finder and on graph algorithms for data processing [1]. The basic idea is to represent the motion of objects in successive images as flows in bipartite graphs. By optimization we get plausible assignments of objects from successive frames. Compared to other approaches which determine correspondences using a greedy nearest neighbor search among the objects' centers of mass the system performs reasonably robust. The robustness has been shown in a multitude of experiments carried out in different environments.

4.1 Finding Objects

To extract single objects two heuristics are applied to the laser scans. The first heuristic assumes that objects form dense clusters in the scan. The second assumption presumes that objects are "almost convex".

To segment a scan sequence into different objects distance gaps are used as split positions for the sequence. Distance gaps occur if there is a significant distance difference between adjacent objects. A threshold value is chosen in advance for the maximum allowed distance gap.

The subsequences of sample points yielded by the preceding step are further divided assuming objects are convex. Our approach is to compute the visible part of the convex hull for each of the given subsequences. For each sample point its distance to the line defined by the next preceding and succeeding sample points on the hull is computed. If there is a point whose distance exceeds a threshold value the sequence is divided at a point with a maximum distance, and this split procedure is applied again in a recursive manner.

4.2 Object Correspondence

As we intend to track objects in a dynamic environment, we have to compare information about objects from successive scans. This is done by a combination of graph algorithms.

From the previous and the current scan two sets of objects $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_m\}$ are given. The goal is to find objects $u_i \in U$ from the previous scan corresponding objects $v_j \in V$ from the current scan. This can be seen as a matching in the bipartite graph $(U \cup V, U \times V)$.

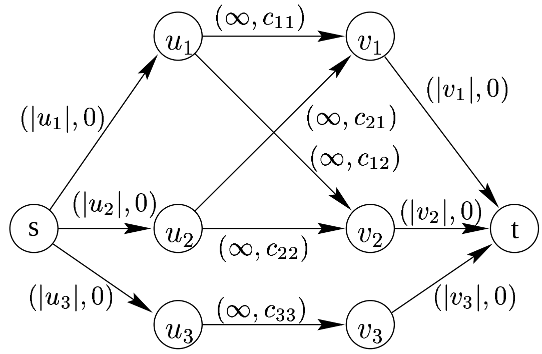


Figure 4: *Graph G for minimum cost maximum flow*

To find a matching representing plausible assignments between objects we start by computing a maximum flow with minimal cost in a graph $G = (\{s, t\} \cup U \cup V, E)$ from source s to sink t as illustrated by figure 4.

Out of all possible matchings for practical reasons the ones are chosen which to do not a threshold value. Finally an object matching is deduced by retaining only edges conveying large amounts of this minimum cost maximum flow, i.e. we compute a maximum weight matching.

5 Extraction of Motion Patterns

When people move in their environment they typically follow certain motion patterns. Their movements are related to specific locations, which attract people or which are avoided and specific trajectories that people are following. The structure of the environment may influence the motion of people. They will have to line up when a narrow doorway is to be passed, they will form up to groups when their destination is the same or they will move all over the place. Recognition and understanding of these motion patterns may lead to a better performance of mobile robots in dynamic environments, especially when the robot interacts with people following or avoiding them.

In this section we propose a statistical method to characterize and recognize typical movement patterns. The fused tracking mechanisms described in earlier sections detect moving and static parts of the environment and estimate trajectories of moving objects. We use this information for

- estimation of crowded areas where people show typical motion behaviors
- sinks and sources of motion streams (where people enter the scene or leave the scene)

- determination of typical movement patterns in the environment

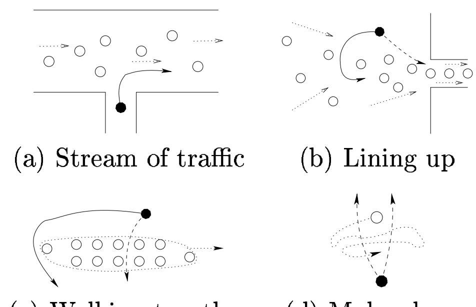
5.1 Recognition of spatio-temporal Sequences

Based on the recognition of simple motion patterns certain dynamic situations can be recognized.

Figure 5a shows the situation where an autonomous mobile robot, depicted as a solid circle, encounters dense traffic when its plan was to perform a right turn. An obviously suboptimal reaction would be to stop and wait for the traffic to decrease. A human would recognize that in this situations it is necessary to carefully but resolutely join the stream, as a decrease or its end is not in sight. Figure 5b shows a situation where a moderate traffic stream is confronted with a narrow passage, so that involved agents have to line up. A human recognizes immediately that he or she has to join the queue at its end in order to get through the narrow passage. But this is a challenging situation for an autonomous mobile robot.

Another example is shown in Figure 5c, where a group of agents is walking together in a highly ordered manner. For instance this might be a basic school class on an excursion. A human recognizes that it would be quite a bad idea to try to join or cross this sort of traffic stream and instead prefers to wait. So this is an example where a robot that is able to recognize and join or cross dense traffic should still behave respectfully.

Last but not least figure 5d shows one human deliberately obstructing the robot. Here a robot that is not aware of this situation surely is trapped.



(a) Stream of traffic (b) Lining up

(c) Walking together (d) Malevolence

Figure 5: *Example situations in crowded environments*

So we expect mobile robots to behave more efficiently in real-world scenarios by increased situation awareness. This seems possible as such robots are able to reason about up to which degree other agents behave respectfully (i.e., decelerate to permit the robot to safely join a traffic stream) or disrespectfully (i.e., deliberately obstruct the robot's desired motion) on the robot's behalf in a certain situation. Recognition of a certain situation may move the robot to adapt its current behavior or choose a new behavior to reflect the newly recognized circumstances.

5.2 Determination of crowded areas

The presented algorithms provide a list of tracked objects T_i which describes the type of object $type_i$ (static or moving), the current position pos_i and the velocity vel_i. This information is stored in several motion maps with different semantics. The absolute motion map $Abs_{x,y}$ stores how much motion occurred in a certain cell $C_{x,y}$:

$$Abs_{x,y} = \sum_i |pos_i| : pos_i \in C_{x,y} \tag{1}$$

The average motion map stores the average motion vector

$$Avg_{x,y} = \sum_i |\frac{vel_i}{Abs_{x,y}}| : pos_i \in C_{x,y} \tag{2}$$

In the variance motion map $Var_{x,y}$ motion vector variances are stored.

Motion areas are estimated from these maps by hierarchical clustering. During initialization a motion cell is created for every position where a minimum motion occurs. The cluster centers are set to the corresponding motion vectors of map $Abs_{x,y}$. Subsequently adjacent cells are merged when their typical motion is similar. Cluster cells are merged until no more similar motion vectors are found.

The hierarchical clustering provides a list of clusters which correspond to areas with similar directed motion. Motion streams are computed on the result of motion cell clustering by analyzing the motion vectors of adjacent clusters. If adjacent cells in adjacent clusters represent similar motion vectors the clusters are put in a common motion stream. Therefore a motion stream consists of adjacent clusters with comparable motion.

Sources and sinks of motion are computed from motion streams simply by analyzing which streams enter or leave the observed scene. If the motion stream at the scene border is narrow we call this area *door*.

To characterize areas with orderless motion, remaining adjacent small clusters which show high motion activity are merged. These cluster cells contain motion vectors which have a high variance either within a single motion cell or between two adjacent cells.

5.3 Simple Motion Patterns

To characterize simple motion patterns we use the fundamental motion elements of tracked objects T_i

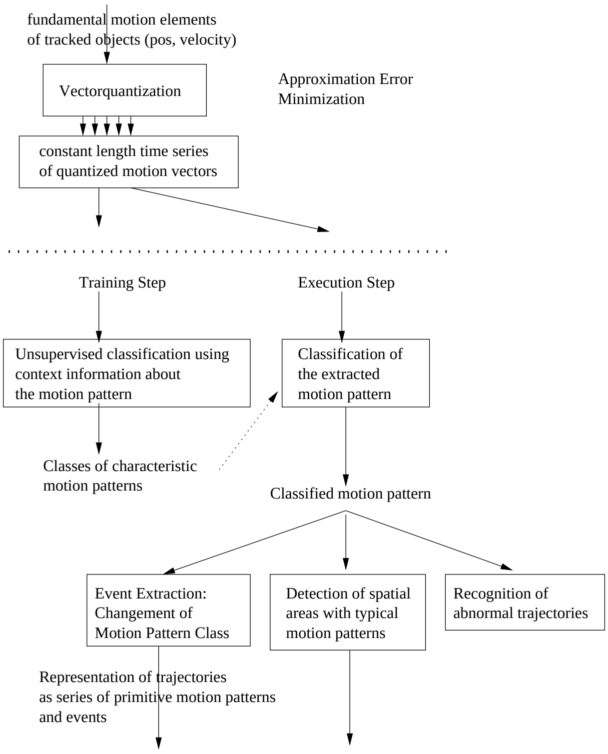


Figure 6: *Extraction of motion patterns*

(figure 6). The most dominant motion vectors are computed by a tree-based vector quantization which minimizes the approximation error. The quantization is applied to constant length trajectories of the tracked persons. Trajectories are represented as histograms of quantized vectors.

During the training step the system acquires representations of typical trajectories. Unsupervised classification which incorporates context information is applied to the total set of time series. The context consists of spatial information estimated during the determination of crowded areas. Motion trajectories are joined into a single cluster when they show a similar motion and their spatial context is the same, fe. door, hallway, area with unordered motion, etc.

During the execution step the characteristic motion patterns are used for classification of a currently extracted motion pattern. Depending on the classification result abnormal trajectories are recognized.

6 Conclusion

The main contribution of this paper are two methods for tracking of people which can be applied to different environments with a moving observer. We treated this problem by integration of range and appearance measuring sensors. While the range information is estimated in a very robust manner by conventional laser range finders, color vision is typically vulnerable to illumination changes. The advantage of color vision is that it allows an easy and robust differentiation between persons by means of color and visual appearance features. Therefore we rely on the robust person detection based on robust range information and use the color for differentiation. Based on the results of the tracking algorithms we implemented a statistical method for recognition of motion patterns and spatial areas where typical motion patterns occur.

Acknowledgments

This work was supported by the German Department for Education and Research (BMB+F) under grant no. 01 IL 902 F6 as part of the project MORPHA.

References

[1] B. Kluge, D. Bank, J. Illmann, C. Köhler, E. Prassler, and M. Strobel. Fast and robust tracking of multiple moving objects with a laser range finder. In *Proc. of the 2001 IEEE Int. Conf. on Robotics and Automation, Seoul*, pages 1683–1688, May 2001.

[2] E. Meier and F. Ade. Object detection and tracking in range image sequences by separation of image features. In *IEEE International Conference on Intelligent Vehicles, Stuttgart, Germany*, pages 176–181, October 1998.

[3] C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, and R. Wörz. Integrating vision-based behaviors with an autonomous robot. *Videre, MIT-Press, ISSN 1089-2788, Volume 1, Number 4, Winter 2000*, 2000.

[4] K. Sobottka and H. Bunke. Vision-based driver assistance using range imagery. In *IEEE International Conference on Intelligent Vehicles, Stuttgart, Germany*, pages 162–167, October 1998.

[5] M. Sullivan, C. Richards, C. Smith, O. Masoud, and N. Papanikolopoulos. Pedestrian tracking from a stationary camera using active deformable models. In *Proc. of Intelligent Vehicles, Detroit*, pages 90–95. IEEE Industrial Electronics Society, September 1995.

[6] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *Photonics East, SPIE, Bellingham, WA.*, 2615, 1995.