Person following and mobile camera localization using particle filters

B. Kwolek
Rzeszów University of Technology, W. Pola 2, 35-959 Rzeszów
bkwolek@prz.rzeszow.pl

Abstract

This paper investigates visual head tracking during person following with a mobile robot. The position of the mobile robot is determined during person following on the basis of laser readings. The tracking of the head is done using a particle filter built on cues such as color, depth, gradient and shape. The appearance of the head is represented by an ellipse with color histogram in its interior and an intensity gradient along the ellipse boundary. The size of the ellipse is computed from a stereo pair. The localization of the moving camera is achieved using a particle filter framework. A particle filter responsible for localization utilizes laser data obtained from sensor and data obtained from a map representing the environment. Current laser readings are compared with scans coming from a global map in respect of histogram intersection. Experimental results, where a person moves facing the cameras, the robot follows the person and simultaneously determines its position, demonstrate the feasibility of the tracking as well as the localization method.

1 Introduction

Service robots are designed for supporting jobs for people in their life environment. They should operate in dynamic and unstructured environment and provide services while interaction with people who are not specially skilled in robot communication. A service robot should be user-independent and ensure a collision free movement for surroundings and the user. Such an intelligent machine should be equipped with a vision system that can ensure an adaptation to the changes in its environment. A vision system can be particularly useful in programming by demonstration of new tasks by non-expert user. For non-expert users it is far easier to point to an object than employ other description of its coordinates.

Robust real-time tracking and segmentation of a moving face in image sequences is a fundamental step in many vision systems. A visual interface may use face tracking and detection techniques to direct the attention of the mobile robot to a human being and maintain the face in the camera's field of view. Thanks to such a reference point a robot can understand simple intentions of the user and in consequence can carry out different useful tasks for people, especially when robot knows its own position in environment.

The localization of a mobile robot consists in determining the position and orientation from an incoming stream of sensor data. Self-localization of a mobile robot in complex environment is a problem of significant importance and difficulty. A mobile agent which knows its position in environment can aid surveillance tasks and provide useful information about human activity. The problem of localization can be divided into two categories, namely global and local. Local localization can be perceived as a part of the global localization and it is often referred to as the pose tracking problem, where the robot knows the starting position and orientation within some certainty and then tries to keep the track of the position while maneuvering. In global localization, which is often referred to as the kidnapped robot problem, the robot should be able to estimate the position without any a priori information about the pose.

A kind of human-machine interaction which is very interesting and has some practical use is following a person with a mobile robot. This behavior can be useful in several applications including robot programming by demonstration and instruction which in particular can rely on tasks comprising a directing a robot to specific place where the user wants point to the object of interest.

The aim of this work was to prepare software that makes possible realization of experiments consisting in tracking of the head to follow a person with a mobile robot as well as simultaneous self-localization of the robot in office environment. A mobile robot Pioneer 2DX [9,14] that was designed to move across a relatively flat surface was used in experiments and tests with the prepared software. The robot was equipped with SRI's Megapixel Stereo Head as well as scanner of SICK. The LMS 200 laser scanner delivers both range and angle information. A typical laptop computer equipped with 2.4 GHz Pentium III is utilized to run the developed software.

In order to escort or to accompany a person, the robot needs to know the relative position of the person. Human faces represent one of the most common patterns and therefore they are perhaps the most useful data source in person detection. The shape of the head is one of the most easily recognizable human parts and can be reasonably well approximated by an ellipse. The discussed later face/head tracking method is based on particle filter combining color, image gradient, depth and shape information, which are used to set the parameters of an ellipse modeling the head on the image plane. During self-localization the robot uses a map previously learned from laser range data. The separate

particle filter is applied to estimate the pose of the robot on-line. A set of samples is used to represent the probability density function encoding the robot knowledge about its position. The particle filter responsible for localization utilizes the histogram intersection to compare the current laser scan with scan representation obtained from the global map of the environment. High similarity of histograms indicates high probability that robot occupies the currently considered pose in the map.

This paper is organized as follows. After discussing related work we will present particle filtering in section III. Then we describe the face tracking algorithm. After that we demonstrate how histogram matching techniques can be used to effectively determine the robot pose in office environment. In section V we present experimental results. Finally, some conclusions are drawn in the last section.

2 Related Work

A lot of work has been performed in the area of human tracking. The Intel CamShift algorithm [2] was designed to handle precise tracking of facial location on the basis of a non-moving camera. It is considered as a very good tracking algorithm especially suited for perceptual interface. Pfinder [13] uses a multi-class statistical model of color and shape to obtain a blob representation of the tracked silhouette in a wide spectrum of viewing conditions. In Birchfield's real-time head tracking system [1], the projection of a head in the image plane was modeled by an ellipse. The intensity gradient near the edge of the ellipse and a color histogram representing the interior were used to update the ellipse parameters over time. Darrell, et al. [3] combine stereo and color via an intensity pattern classification method to track people. The original application of the particle filter in computer vision was for object tracking in an image sequence [8]. Global color reference models and Bhattacharyya coefficient as a similarity measure between the color distribution of the model and target candidates have been used in a Monte Carlo tracker [11]. Fox, Burgard et. al. [6,4] introduced a family of Monte Carlo based algorithms, called Monte Carlo Localization. The algorithms are widely used to solve many different forms of localization, including global localization, position tracking.

3 Overview of Particle Filtering

The particle filter is an algorithm for estimating the posterior state of a dynamic system over time where the state cannot be measured directly, but may be estimated at the current time-step *t*, given the initial state, all sensor measurements $z^t = z_0,...,z_t$ and controls $u^t = u_0,...,u_t$ up to the current time. The particle filter computes the posterior recursively using the Bayes filter equation

$$p(x_t \mid z^t) = \eta \; p(z_t \mid x_t) \int p(x_t \mid x_{t-1}, u_{t-1}) p(x_{t-1} \mid z^{t-1}) dx_{t-1}$$

where η is a normalization constant. To implement this recursive equation one needs to know initial condition $p(x_0 \mid z^0)$, the next state probabilities $p(x_t \mid x_{t-1}, u_{t-1})$ and the observation likelihood $p(z_t \mid x_t)$. The current state x_t is only dependent on previous state x_{t-1} and a known control input u_{t-1} according to the probabilistic action model $p(x_t \mid x_{t-1}, u_{t-1})$. The measurement z_t is conditionally independent of earlier measurements z^{t-1} given x_t. The perceptual model $p(z_t \mid x_t)$ describes the probability for taking certain measurements at certain locations. It depends on the type of sensor being used and takes into account the noise that appears in the sensor readings. At the initial time step, without prior information we assume that the initial state is uniform over all allowable states. In case of tracking the initial position is typically specified through Gaussian centered around x_0. In visual tracking, for example, x_t can represent the position and orientation of the human face. In mobile robot localization a three-dimensional state vector $[\mathrm{x}, \mathrm{y}, \theta]^T$ representing the position and rotational heading direction of the robot is used typically.

In a particle filter the current state of the target is modeled as the density of a set of particles. The particles provide a mechanism for maintaining multiple hypotheses and propagating the uncertainty over time. A large enough set of weighted particles can reflect the true posterior density. The idea of the particle filter is to represent the posterior by a set S_t of *N* weighted particles distributed according to posterior: $p\left(x_t \mid z^t\right) \approx \left\{x_t^{(i)}, w_t^{(i)}\right\}_{i=1,...,N}$, where $x_t^{(i)}$ is a particle and $w_t^{(i)}$ are non-negative weights called importance factors, which sum up to one.

The probabilistic search for the best state is realized on the basis of motion as well as observation model of the particle filter. The particle filter operates thus in two alternating phases: prediction and update. First all particles are moved according to motion model. In the update phase sensor information is incorporated into probability distribution. The update is done by multiplying the weight $w_t^{(i)}$ of each sample $x_t^{(i)}$ by the probability of observing z_t at the position given by $x_t^{(i)}$. Particle filters work well when the conditional densities $p(z_t \mid x_t)$ are reasonably flat.

A crucial step in the particle filter is re-sampling. The aim of the re-sampling which has been introduced by Gordon et. al. [7] is to eliminate particles with low importance weights and multiply particles with high importance weights. The re-sampling selects with higher probability samples that have a high likelihood associated with them, while preserving the asymptotic approximation of the sample based posterior representation. Without re-sampling the variance of the weight increases stochastically over time [5].

The sensor model $p(z_t \mid x_t)$ describes how likely it is to obtain a particular sensor reading z_t given state x_t. This probability is often computed by estimating the sensor

reading $\tilde{z}_t$ in state x_t and determining some distance $dist(z_t, \tilde{z}_t)$ between the given sensor reading z_t and the estimation $\tilde{z}_t$ resulting from the model. This distance is then mapped to a probability.

During head tracking the head has been modeled in the 2D image domain by an ellipse. The color distribution within interior of the ellipse is represented by a color histogram. The color histogram is dynamically updated over time. The lengths of the ellipse's minor axis are determined on the basis of depth information. The particles representing the candidate ellipses are weighted in each time step in respect of intensity gradient near the edge of the ellipse and matching score of the color histograms representing the interior of the ellipse surrounding the tracked object and currently analyzed one during the update stage.

In mobile robot localization task each particle can be seen as the hypothesis of the robot being located at particular position. The sample weight represents the likelihood of a particular sample being the true target pose and is calculated by comparing the sensor data to data obtained from the prepared in advance map of the environment. Our localization approach focuses on histogram based techniques to compare the current laser scan with the scan representation obtained from an existing map. A high similarity of histograms indicates a good match between laser readings and scans representing considered map pose. The histogram based map representation has powerful capability and can be used to distinguish sensor scans in a very fast manner.

Particle filters are attractive in robotics for several reasons. They utilize imperfect perception models and incorporate imperfect sensor data through Bayes rule. The ability to represent multimodal posterior densities allows them to globally localize as well as relocalize the robot in case of failure. Algorithms that deal with the global localization are relatively recent, although the idea of estimating state recursively using particles is not new.

4 Head Tracking Using Particles

The shape of the head is one of the most easily recognizable human parts and can be reasonably well approximated by an ellipse. In our approach, an ellipse based head likelihood model, consisting of gradient along the head boundary as well as a matching score between color histograms as a representation of the interior of (i) an ellipse surrounding the tracked object and (ii) a currently considered ellipse, together with depth information is utilized to find the weights of particles during tracking. Particle locations where the weights have large values are then considered to be the most likely locations of the object of interest.

In order to obtain information about possible location of the tracked target we use color histogram matching techniques. The main idea of such an approach is to compute color distribution at the hypothesized region in form of the color histogram from the ellipse's interior and to compare it with the computed in the same manner histogram representing the tracked object in the previous iteration. The smaller the discrepancy between the candidate histogram representing the ellipse's interior at specific particle position and the reference histogram from previous iteration, the higher the probability that the tracked target is located inside the candidate region. The outcome of the histogram matching that is combined with gradient information is used to provide information about expected target location and is utilized during weighting particles.

In the context of head tracking on the basis of images coming from a mobile camera the features which are invariant under head orientations are particularly useful. In general, histograms are invariant to translation and rotation of the object and they vary slowly with the change of angle of view and with the change in scale. A histogram is obtained by quantizing the ellipse's interior colors into K bins and counting the number of times each discrete color occurs. Due to the statistical nature, a color histogram can only reflect the content of images in a limited way and thus the contents of the interior of the ellipses taken at small distances apart are strongly correlated. If the number of bins K is to high, the histogram is noisy. If K is too low, density structure of the image representing the ellipse's interior is smoothed. Histogram based techniques are effective only when K can be kept relatively low and where sufficient data amounts are available. The reduction of bins makes a comparison between the histogram representing the tracked head and the histogram of candidate head faster. Additionally, such a compact representation is tolerant to noise that can result from imperfect ellipse-approximation of a highly deformable structure and curved surface of a face causing significant variations of the observed colors.

Color information is particularly useful to support a detection of faces in image sequences because of robustness towards changes in orientation and scaling of appearance of object being in movement. The efficiency of color segmentation techniques is especially worth to emphasize when a considered object is occluded during tracking or is in shadow. Skin colors acquired from a static person tend to form tight clusters in several color spaces while colors acquired from a moving person form widen clusters due to seeming changes in reflecting surfaces. To make the histogram representation of the tracked head less sensitive to lighting conditions the HSV color space has been chosen and the V component has been represented by 4 bins while the HS components obtained the 8-bins representation.

In order to compare histograms we have implemented the histogram intersection technique [12]. For a given pair of histograms I and M, each containing j values, the intersection of the histograms is defined as follows: $H = \sum_{u=1}^{K} \min(I^{(u)}, M^{(u)})$. The terms $I^{(u)}$, $M^{(u)}$ represent the number of pixels inside the u-th bucket of the candidate histogram and the histogram representing the tracked head, respectively, whereas K the total number of buckets. The result of the intersection of two histograms is the number of pixels that have the same color in both

histograms. To obtain a match value between zero and one the intersection is normalized and the match value is determined as follows: $H_{\cap} = H / \sum_{u=1}^{K} I^{(u)}$.

The length of the minor axis of a considered ellipse is determined on the basis of depth information. Taking into account the length of the minor axis resulting from the depth information we also considered smaller and larger projection scale of the ellipse and therefore a larger as well as smaller minor axis about one pixel have been taken into account as well. The length of the minor axis has been maintained by performing the local search to maximize the goodness of the following match: $q^{*} = \arg\max_{q_i \in Q} \{G(q_i) + H_{\cap}(q_i)\}$, where G and $H_{\cap}$ are normalized scores based on intensity gradients and color histogram intersection. Particularly, if the length of minor axis of the considered ellipse was different from the length of minor axis of the reference ellipse representing the tracked head, in order to provide *j* values in the histogram *I* a histogram normalization with respect to ellipse's area has been realized. The search space Q comprises the ellipse's length obtained on the basis of depth information as well as smaller/larger minor axes about one pixel.

The discussed method of target representation has a construction phase and a run phase. In the construction phase which is realized of-line the elliptical upright outlines as well as masks containing interior pixels have been prepared and stored for the future use. We have assumed that a reference ellipse is located in a central point in a candidate region. Such a candidate area considers all expected head locations which can occur in the next time step. We have then fixed a search strategy allowing us to compute histogram iteratively, i.e. considering adjoining ellipses when processing from top to bottom and from left to right and then from right to left, etc. For each location in the assumed candidate area we constructed a list of positions which should be substituted (added and removed) in the current histogram to determine the histogram at the next location in the utilized search strategy. As a result, for each possible length of the minor axis we obtained a fast strategy to match histograms representing hypothetical head locations in the candidate area. In an on-line phase this strategy allows us to compute the likelihood of each candidate head location and store this information in a two dimensional table, which can be easily accessed during weighting of samples.

The histogram representing the tracked head has been adapted over time. This makes possible to track not only a face profile which has been shot during initialization of the tracker but in addition different profiles of the face as well as the head can be tracked. The actualization of the histogram has been realized on the basis of the equation $M_t^{(u)} = (1-\alpha) M_{t-1}^{(u)} + \alpha I_t^{(u)}$, where I_t represents the histogram of the best-fit ellipse interior, whereas $u = 1...K$.

The weight $w_t^{(i)}$ of each hypothetical head region is dependent on normalized intensity gradients and color histogram intersection which were obtained in the local search in the space Q for the length of the minor axis. We currently use a first order motion model describing a region which moves with constant velocity. The samples are propagated on the basis of a dynamic model $s_t = A s_{t-1} + v_t$, where A denotes a deterministic component describing a constant velocity movement and v_t is a multivariate Gaussian random variable. The diffusion component represents uncertainty in prediction and therefore provides a way of performing a local search about the state for the best-fit ellipse.

5 Robot Localization Using Particles and Histogram Matching

In our localization approach the sensor model describes the probability of obtaining a particular scan shape given the laser's pose and a geometrical map of the environment. This probability is computed by estimating the sensor reading $\tilde{z}_t$ at pose x_t and determining distance $dist(z_t, \tilde{z}_t)$ between the given sensor reading z_t and the estimation $\tilde{z}_t$ resulting from the geometrical model of environment. The distance is determined through histogram matching and then mapped to a probability.

A single scan of the laser range finder which was used in experiments returns a hemicircle of 180 readings with 1 degree incrementation. The range error of the laser is 1 cm. A sample laser scan is depicted in the Fig. 1a. A reference scan which has been obtained on the basis of the map for corresponding robot pose is indicated in the Fig. 1b. Fig. 3. illustrates the geometrical map of environment in which localization experiments have been conducted. This office-like environment is 560 by 460 cm and it has been discretized into 280x230x90 cells.

In order to predict the probability distribution representing the pose of the mobile robot we need an action model. Any arbitrary mobile robot motion $[\Delta x, \Delta y]^T$ can be achieved as a rotation that sets the robot heading towards the target location, followed by a translation that moves the robot to the target position. The noise is applied separately to each of the two types of motion because they are independent. When the robot rotates about $\Delta\theta$ the odometry noise can be modeled as a Gaussian with experimentally established mean *m* and standard deviation σ_{rot} proportional to $\Delta\theta$. The orientation of each particle is updated by adding $\Delta\theta$ and random number drawn from normal distribution $N(m, \frac{\Delta\theta}{360}\sigma_{rot})$ as follows: $\theta_t = \theta_{t-1} + \Delta\theta + N(m, \frac{\Delta\theta}{360}\sigma_{rot})$. Modeling the forward translation of the mobile robot is more complicated. The first error is related to distance that has been traveled and the second is associated with changes of the orientation attending forward translation. The simple

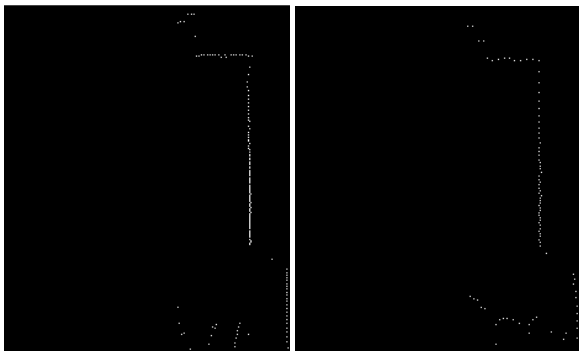

Fig. 1. Laser data from environment. Data from map

way to obtain translation model is to discretize the distance into J equal steps and to cumulate the simulated effect of noise from each step. Taking into account the kinematics of our robot this can be achieved in the following manner:

$$\begin{pmatrix} x_t \\ y_t \\ \theta_t \end{pmatrix} = \begin{pmatrix} x_{t-1} + (\Delta\rho + \varepsilon_{\Delta\rho})\cos(\theta_{t-1} + \varepsilon_{\Delta\theta}) \\ y_{t-1} + (\Delta\rho + \varepsilon_{\Delta\rho})\sin(\theta_{t-1} + \varepsilon_{\Delta\theta}) \\ \theta_{t-1} + \varepsilon_{\Delta\theta} \end{pmatrix}$$

where $\varepsilon_{\Delta\rho} = N(0,1)\sigma_{tr}\sqrt{J}\Delta\rho$, $\varepsilon_{\Delta\theta} = N(0,1)\sigma_{dr}\sqrt{J}\Delta\rho$, $N(0,1)$ is a random number drawn from a Gaussian distribution, σ_{tr} and σ_{dr} are experimentally obtained values per 1 m distance traveled. Fig. 2. demonstrates the probability distribution of robot location after translation of 100 cm and 200 cm, respectively, for σ_{tr} =1cm/m and σ_{dr} =5°/m.

In order to obtain an estimate of the robot pose the weighted mean ($\sum_i w_i x_i$), in a small sub-cube around the best particle has been utilized. The orientation of the robot has been determined on the basis of sum of direction vectors of particles from sub-cube as follows:

$$\theta = \text{atan2}\left(\sum_i \sin\theta_i, \sum_i \cos\theta_i\right)$$

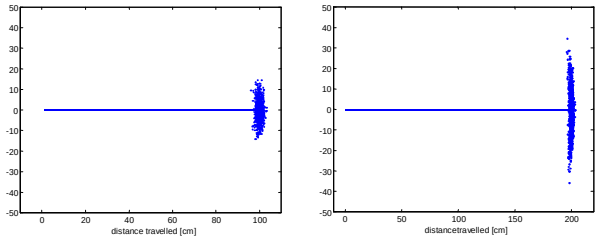

Fig. 2. Motion model representing the uncertainty

6 Experiments

All experiments were carried out in an office environment with our experimental Pioneer 2DX based platform which is equipped with a laser range finder, color stereo pair as well as on board 2.4 GHz laptop computer. Two 4-bins histograms representing x and y-components of the laser scans have been used in the particle filter responsible for localization. In order to evaluate the precision of determining the position in certain points, we conducted experiments in an office environment and utilized the map shown in the Fig. 3. The average error between the goal position and the position reported by robot was 10 cm. Fig. 3. demonstrates the localization stage after 8 and 15 iteration. We can observe that in the 8 iteration the robot knows its position. The mentioned above results have been obtained on the basis of 1000 particles. The effect of probabilistic search for the best position has been amplified via a local change in the position of particles according to their probability. The more probable the particle was, the less it was moved. Such an operation acknowledged his particular usefulness in the global re-localization of the robot.

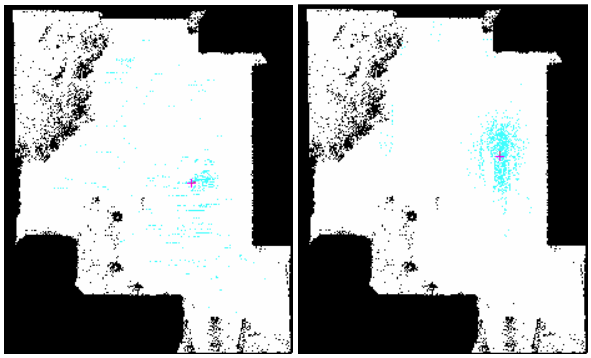

Fig. 3. Global localization of the mobile robot

The tracking algorithm has been tested by controlling the robot equipped with the stereo pair. At the beginning of tests we realized experiments consisting in a rotation of mobile robot. In such experiments a user moved about a room, walked back and forth as well as around the mobile robot. The aim of such a scenario was to evaluate the quality of ellipse scaling in response of varying distance between the camera and the user. The aim of the robot orientation controller is to keep the position of the tracked face at specific position in the image. Our experimental results show that thanks to stereovision the ellipse is properly scaled and sudden changes of the minor axis length as well as ellipse's jumps are considerably eliminated, having on regard a version of the algorithm with no depth information. The tracking algorithm was implemented on the robot on board computer, see Fig. 4, and runs at frame rates about 10 Hz depending on image complexity. Selected frames from a tracking sequence are presented in the Fig. 5, where the wooden door is close in color space to the face being tracked.

Following a person with mobile robot is a much more challenging task than ones with a fixed camera because of both motion of the camera and of the user [10]. Once the face of the user is located, the controller using the data which are provided by the tracker keeps the head within the camera field of view through steering the robot. The aim of the robot orientation controller is to keep the position of the

tracked face at specific position in the image plane. The distance of 1.6 m between the camera and the user has been assumed as the reference value that the linear velocity controller should maintain during person following. To eliminate needless robot rotations as well as forward and backward movements we have applied a simple logic

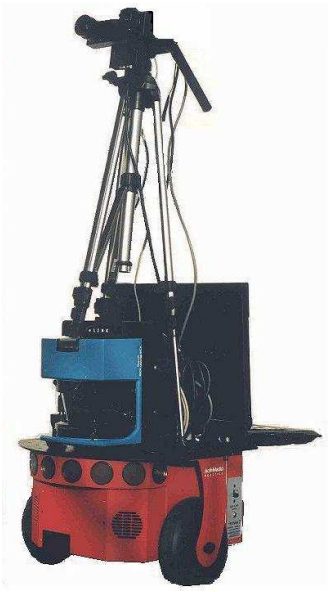

Fig. 4. The robot

providing necessary insensitivity zone in utilized PD controllers. The presented approach enables the robot to follow a person at speed up to 25 cm per second during simultaneous localization. The stereo pair has been turned about 180^{o} regarding orientation presented in the Fig. 4.

Fig. 5. Face tracking (frame #1 and #500)

7 Conclusions

We have presented a system that robustly tracks the face of the user and simultaneously determines the position of the mobile robot. To show the feasibility of the system, we have conducted several experiments with a real robot. A histogram based representation of the environment is very useful in particle filter based mobile robot localization. Experimental results, where a person moves facing the cameras and robot follows the person and determines its position demonstrate the feasibility of the tracking as well as localization method.

Acknowledgment

This work has been supported by the Polish Committee for Scientific Research (KBN) within the project 4T 11C 01224

References

1. Birchfield S.: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, IEEE Conf. on Computer Vision and Pattern Rec., Santa Barbara, 1998, 232-237
2. Bradski G. R.: Computer Vision Face Tracking as a Component of a Perceptual User Interface, In Workshop on Applications of Computer Vision, Princeton, 1998, 214-219
3. Darrell T., Gordon G., Harville M., Woodfill J.: Integrated Person Tracking Using Stereo, Color, and Pattern Detection, Proc. of the Conf. on Computer Vision and Pattern Recognition, Santa Barbara, 1998, 601-609
4. Dellaert F., Fox D., Burgard W., Thrun S.: Monte carlo localization for mobile robots. In Proc. of the IEEE Int. Conf. on Robotics and Automation, 1999, 1322-1328
5. Doucet A., Godsill S., Andrieu Ch.: On sequential Monte Carlo sampling methods for Bayesian filtering, Statistics and Computing, vol. 10, 2000, 197-208
6. Fox D., Burgard W., Dellaert F., Thrun S.: Monte carlo localization: Efficient position estimation for mobile robots. In Proc. of the Sixteenth National Conference on Artificial Intelligence, Orlando, FL, 1999, 343-349
7. Gordon N, Salmond D, Smith A.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEEE Trans. Radar, Signal Processing, vol. 140, 1993, 107-113
8. Isard M., Blake A.: Contour Tracking by Stochastic Propagation of Conditional Density, European Conf. on Computer Vision, Cambridge, 1996, 343-356
9. Kortenkamp D., Bonasso R. P., Murphy R. (Ed.): Artificial intelligence and mobile robots-Case studies of successful robot systems, The MIT Press, Cambridge, London, 1998
10. Kwolek B.: Face Tracking System Based on Color, Stereovision and Elliptical Shape Features, Proc. of the IEEE Conf. on Advanced Video and Signal Based Surveillance, Miami, FL, IEEE Comp. Society, 2003, 21-26
11. Perez P., Hue C., Vermaak J., Gangnet M.: Color-Based Probabilistic Tracking, European Conference on Computer Vision, 2002, 661-675
12. Swain M. J., Ballard D. H.: Color indexing, Int. J. of Computer Vision, vol. 7, no. 1, 1991, 11-32
13. Wren C., Azarbayejani A., Darrell T., Pentland A.: Pfinder: Real-Time Tracking of the Human Body, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 1997, 780-785
14. Pioneer 2 mobile robots, ActivMedia Robotics, 2001