Perception Issues in Mobile Robotics

Juan Andrade Cetto

Institut de Robòtica i Informàtica Industrial
Universitat Politècnica de Catalunya - CSIC
Llorens Artigas 4-6, Barcelona 08028, Spain
cetto@iri.upc.es

Abstract. This paper presents some recent advances in perception for mobile robotics The main objective is, by fusing multisensor information, to obtain sufficiently robust perception capabilities for the successful completion of typical mobile robotics tasks, including human-machine interaction. The use of these techniques is tailored at service robotics applications, such as cleaning of large industrial areas, inspection, security, transportation, costumer service, etc.

The article focuses in two specific topics: a) simultaneous localization and mapping of large indoor industrial settings, and b) localization and tracking of human faces in image sequences. We tackle the simultaneous localization and mapping problem (SLAM) from an stochastic estimation perspective, and propose techniques to solve the effects of partial observability inherent with having two coupled inference problems. With respect to the localization and tracking of human faces (and other objects), we have developed an articulated head mounted vision system capable of tracking an individual under the most severe changing illumination conditions. With this platform, we have tested a mirage of estimation techniques for tracking color histograms.

1 Introduction

During the last few years, our efforts at the Learning and Vision Mobile Robotics Group have been tailored at giving our mobile platforms the ability to navigate autonomously in unknown structured settings. In this sense, we have contributed new insight in the classical simultaneous localization and map building problem, from a control systems theory point of view [1–5]. Furthermore, we have developed new feature validation techniques that improve the robustness of typical map building algorithms [6, 7].

Moreover, very good results have been achieved in the tracking of subjects under varying illumination conditions and in cluttered scenes. On the one hand, we have mastered now the use of histogram based techniques to color segmentation and illumination normalization [8–11]. On the other hand, we have tested different statistical estimation paradigms to track subject candidates using not only color information but shape as well [12, 13]. Most of our video demonstrations from last year show results in this topic.

As a result of these efforts, within the past three years, the mobile robotics platforms developed in our group have been portrayed numerous times on live and printed me-

5to Congreso Internacional en Control, Instrumentación Virtual y Sistemas Digitales
Mexico, Septiembre 2004. In Avances en Control Digital, Instumentación Virtual, Sistemas Digitales, Arquitectura de Computadoras, Robótica y Procesamiento de Señales. S. Suárez, C. Aguilar, and J. Figueroa Eds., Resaearch on Computer Sciene Series, Instituto Politécnico Nacional. ISBN 970-36-0180-4.

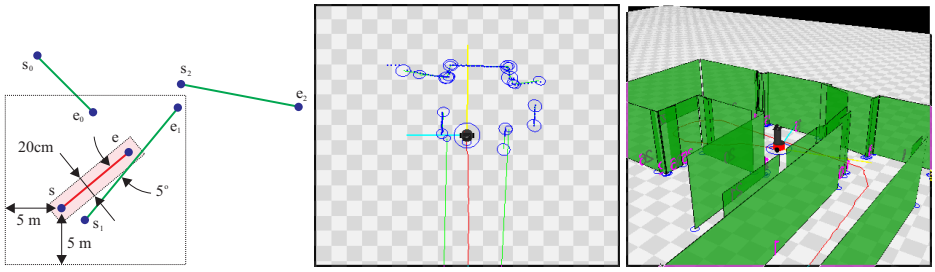


Fig. 1. Simultaneous mobile robot localization and map building. a) Hypothesis search range for walls extracted from a laser range scan. b) The blue dots indicate sensor raw data coming from a laser range finder. The green lines represent walls inferred from consecutive readings. The red lines indicate the estimated robot trajectory. c) Graphical representation of the map built.

dia[1]. Financial support comes mainly from a continued set of projects from the Spanish Interministerial Council of Science and Technology (CICYT)[2].

In the following sections I summarize the key contributions on the aforementioned two topics: simultaneous localization and mapping, and face and object tracking under varying illumination conditions. Each of them tackling very different problems typically encountered in mobile robotics applications.

2 Simultaneous Localization and Map Building

2.1 Iterative End Point Fit

To univocally identify landmarks from sensor data, we study the mathematical foundation necessary to extract the features that build them from laser range data. The features extracted from just one sensor may not suffice in the invariant characterization of landmarks and objects, pushing for the combination of information from multiple sources.

An efficient algorithm for the extraction of straight line segments from noisy data, originally by Douglas and Peucker [14], was conceived primarily to reduce the number of points required to represent a vector encoded polygonal digitized line. Within the cartography community, this problem is referred as the *line simplification problem*. Douglas and Peucker's approach to line simplification is probably the most cited algorithm when it refers to cartographic generalization. We also find the technique nicely applicable to the extraction of straight landmarks from laser and sonar data in mobile robot indoor navigation, as is the case with walls and furniture.

[1] La Politècnica, Num 4. UPC: Criatures Cibernètiques (2001), Canal Blau TV: The IRI Robots in Canal Blau (2002), El Periòdico de Catalunya: El Robot Doméstico Llama a la Puerta (2002), Televisión Española: Informe Semanal (2002), Televisió de Catalunya: Robot Casolà (2003), UPC Libraries: More Than Just Machines (2003).

[2] TAP96-0629-C04-03, TAP98-0473, and DPI2001-2223

A set of ordered points in a plane, $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$, forms a polygonal chain of line segments $\overline{\mathbf{v}_0\mathbf{v}_1}, \overline{\mathbf{v}_1\mathbf{v}_2}, \ldots, \overline{\mathbf{v}_{n-1}\mathbf{v}_n}$. Given a chain C with n segments, the line simplification problem asks for the best chain C', with fewer segments that represents C *well*. Furthermore, we assume that the chain C is simple, that is, C has no self-intersections.

Representing *well* has many possible meanings. For example, that C and C' are close to each other, that the area between C and C' be small, or that other measures of curve discrepancy be small. A recursive version of the original algorithm is shown in Table 1. The key point of the algorithm resides in the technique used to find the appropriate point where to subdivide a line into consecutive line segments. In the original method, subdivision takes place at the farthest point in C from $\overline{\mathbf{v}_i\mathbf{v}_j}$. Table 2 shows the algorithm to find the farthest point to a line in a stream of planar data points. The notation $\tilde{\mathbf{v}}$ indicates homogeneous coordinates, i.e., three dimensional points with the last dimension equal to one.

The original algorithm has quadratic worst-case time complexity. However, a revised version of the algorithm by Hershberger and Snoeyink [15] that uses path hulls and the geometric structure of the problem allows for a reduction of the time complexity of the algorithm to $O(n \log n)$. In our map building implementation [16] we have opted for this latter version. The reader is referred to the proceedings from the talk by Hershberger and Snoeyink [15] for the details of the algorithm. Fig. 1b shows the results of line simplification from raw depth data with a Leuze Rotoscan RS4 and our mobile robot Marco.

```
LineSimplification(C, i, j)
      FindSplit(C, i, j)
      if d > ε
            LineSimplification(C, i, f)
            LineSimplification(C, f, j)
      else
            return v_i v_j
```

Table 1. Recursive version of the Douglas and Peucker algorithm to line simplification.

2.2 Extended Kalman Filter Approach to SLAM

Once landmarks are accurately extracted and identified, the second part of the problem is to use these observations for the localization of the robot, as well as the refinement of the landmark location estimates. We consider robot motion and sensor observations as stochastic processes, and treat the problem from an estimation theoretic point of view, dealing with noise by using probabilistic methods.

Fig. 1a shows some of the model compatibility heuristics devised for the validation of straight lines extracted from laser range data into walls. Frame b shows data as ex-

```
FindSplit(C, i, j)
    d = 0
    for k = i + 1 to j − 1
        dk = |ṽ_i ṽ_j ṽ_k| / ||v_i − v_j||^2
        if dk > d
            d = dk
            f = k
    return f, d
```

Table 2. Algorithm to find the vertex $\mathbf{v}_k$ farthest from the line $\overline{\mathbf{v}_i\mathbf{v}_j}$, and the squared distance to it.

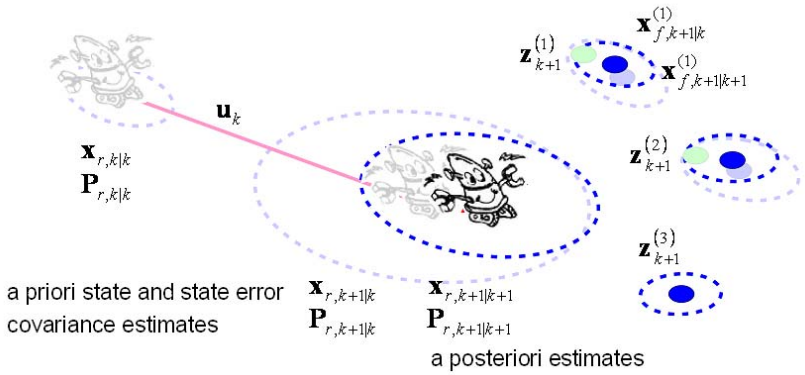


Fig. 2. State estimation approach to simultaneous localization and map building.

tracted from a laser range finder, and 2σ covariance ellipses around hypothesized landmark estimates. The third frame shows a virtual reality model of the map constructed during a run of the algorithm.

The study of stochastic models for simultaneous localization and map building in mobile robotics has been an active research topic for over fifteen years. Within the Kalman filter (KF) approach to SLAM, seminal work by Smith and Cheeseman [17] suggested that as successive landmark observations take place, the correlation between the estimates of the location of such landmarks in a map grows continuously. This observation was ratified by Dissanayake *et al.* [18] with a proof showing that the estimated map converges monotonically to a relative map with zero uncertainty. They also showed how the absolute accuracy of the map reaches a lower bound defined only by the initial vehicle uncertainty.

In spite of these fundamental convergence properties of the KF approach to SLAM, there exist some limitations that still hinder full development of SLAM applications.

The three most criticized of these limitations are the time and space complexity of the algorithm, the restriction to unimodal zero mean white Gaussian models of uncertainty, and the data association and landmark quality problems. We go even further in our analysis of the SLAM problem, and show one more fundamental limitation of the algorithm, that full reconstruction of the map state vector is not possible with a coupled measurement model, regardless of the vehicle model chosen, and that the expected error in state estimation is proportional to the number of landmarks used.

An explicit solution to the SLAM problem for a one-dimensional vehicle called the *monobot* was presented by Gibbens *et al.* [19]. It shed some light on the relation between the total number of landmarks and the asymptotic values for the state error covariance $\mathbf{P}$. They observed for example, that in SLAM, the rate of convergence of $\mathbf{P}$ is fixed, and that its asymptotic value is independent of the plant variance. In their solution to the 1-d Brownian motion SLAM case, the state error covariance is linked to the total number of landmarks in the form of the total Fisher information $I_T = \sum_1^n (1/\sigma_w^2)$. The expression indicates the "*informational equivalence of the measurements and the innovations*" [20], and was derived from a simple likelihood function, one that does not contain the fully correlated characteristics of the measurement model. We derive a more general expression for the total Fisher information in SLAM that shows explicitly the unobservable directions of the map state.

In summary; in SLAM, the state space constructed by appending the robot pose and the landmark locations is fully correlated; a situation that hinders full observability. Moreover, the modelling of map states as static landmarks yields a partially controllable state vector. The identification of these problems, and the steps taken to palliate them, are covered in this section.

Formally speaking, the motion of the robot and the measurement of the map features are governed by the discrete-time state transition model

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{w}_k) \tag{2}$$

The state vector $\mathbf{x}_k^\top = [\mathbf{x}_{r,k}^\top, \mathbf{x}_f^\top]^\top$ contains the pose of the robot $\mathbf{x}_{r,k}$ at time step k, and a vector of stationary map features $\mathbf{x}_f$. The input vector $\mathbf{u}_k$ is the vehicle control command, and $\mathbf{v}_k$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{V}$, representing unmodeled robot dynamics and system noise of the possibly nonlinear difference equation $\mathbf{f}$. Both, the inaccuracies of the also possibly nonlinear observation model $\mathbf{h}$, and the measurement noise, are represented by the zero mean Gaussian random vector $\mathbf{w}_k$, with covariance matrix $\mathbf{W}$. See Fig. 2.

Provided the set of observations $Z^k = \{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ was available for the computation of the current map estimate $\mathbf{x}_{k|k}$, the expression

$$\mathbf{x}_{k+1|k} = \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0}) \tag{3}$$

gives an a priori noise-free estimate of the new locations of the robot and map features after the vehicle control command $\mathbf{u}_k$ is input to the system. Similarly,

$$\mathbf{z}_{k+1|k} = \mathbf{h}(\mathbf{x}_{k+1|k}, \mathbf{0}) \tag{4}$$

constitutes a noise-free a priori estimate of sensor measurements.

Given that the landmarks are considered stationary, their a priori estimate is simply $\mathbf{x}_{f,k+1|k} = \mathbf{x}_{f,k|k}$; and the a priori estimate of the map state error covariance showing the increase in robot and landmark localization uncertainty is given by

$$\mathbf{P}_{k+1|k} = E[\tilde{\mathbf{x}}_{k+1|k}\,\tilde{\mathbf{x}}_{k+1|k}^{\top}] = \mathbf{F}_{\mathbf{x}}\mathbf{P}_{k|k}\mathbf{F}_{\mathbf{x}}^{\top} + \mathbf{F}_{\mathbf{v}}\mathbf{V}\mathbf{F}_{\mathbf{v}}^{\top} \tag{5}$$

The Jacobian matrices $\mathbf{F}_{\mathbf{x}}$ and $\mathbf{F}_{\mathbf{v}}$ contain the partial derivatives of $\mathbf{f}$ with respect to $\mathbf{x}$ and $\mathbf{v}$, evaluated at $(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0})$. Assuming that a new set of landmark observations $\mathbf{z}_{k+1}$ coming from sensor data has been correctly matched to their map counterparts, one can compute the error between the measurements and the estimates with $\tilde{\mathbf{z}}_{k+1|k} = \mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}$.

This error aids in revising the map and robot locations. The a posteriori state estimate is

$$\mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \mathbf{K}\tilde{\mathbf{z}}_{k+1|k} \tag{6}$$

and the Kalman gain is computed with

$$\mathbf{K} = \mathbf{P}_{k+1|k}{\mathbf{H}_{\mathbf{x}}}^{\top}\mathbf{S}^{-1} \tag{7}$$

where $\mathbf{S}$ is termed the measurement innovation matrix,

$$\mathbf{S} = \mathbf{H}_{\mathbf{x}}\mathbf{P}_{k+1|k}{\mathbf{H}_{\mathbf{x}}}^{\top} + \mathbf{H}_{\mathbf{w}}\mathbf{W}{\mathbf{H}_{\mathbf{w}}}^{\top} \tag{8}$$

and $\mathbf{H}_{\mathbf{x}}$ and $\mathbf{H}_{\mathbf{w}}$ contain the partial derivatives of $\mathbf{h}$ with respect to $\mathbf{x}$ and $\mathbf{w}$, and evaluated at $(\mathbf{x}_{k+1|k}, \mathbf{0})$.

Finally, the a posteriori estimate of the map state error covariance must also be revised once a measurement has taken place. It is revised with the Joseph form to guarantee positive semi-definiteness.

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}\mathbf{H}_{\mathbf{x}})\,\mathbf{P}_{k+1|k}\,(\mathbf{I} - \mathbf{K}\mathbf{H}_{\mathbf{x}})^{\top} + \mathbf{K}\mathbf{W}\mathbf{K}^{\top} \tag{9}$$

The contribution to the revision of the robot pose and landmark location estimates is proportional to our degree of trust in the motion and sensor models, respectively. If the plant error covariance $\mathbf{V}$ is large, and the measurement error covariance $\mathbf{W}$ is small, the EKF-SLAM algorithm trusts the observations more than dead-reckoning, revising more heavily the robot pose estimate than that of the landmarks. Conversely, when the measurement error covariance is larger than the plant error covariance, the algorithm trusts more on the motion of the robot and ends up revising more heavily the landmark estimates.

Partial observability of the system (1-2) hinders full reconstructibility of the state space, making the final map estimate dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix. Partial controllability on the other hand, makes the filter believe after a small number of iterations, that it has accurate estimates of the landmark states, with their corresponding Kalman gains converging to zero. That is, after a few steps, innovations are useless. We will next show the unobservable state space directions in the Fisher Information Matrix, and show how

to palliate the effects of full correlation and partial controllability by modifying the measurement model in SLAM.

Under the Gaussian assumption for the vehicle and sensor noises, the Kalman filter is the optimal minimum mean square error estimator. And, as pointed out in [20], minimizing the least squares criteria $E[\tilde{\mathbf{x}}_{k+1|k+1}\tilde{\mathbf{x}}_{k+1|k+1}^{\top}]$, is equivalent to the maximization of a likelihood function $\Lambda(\mathbf{x})$ given the set of observations Z^k; that is, the maximization of the joint probability density function of the entire history of observations, conditioned on the state $\mathbf{x}$

$$\Lambda(\mathbf{x}) = \prod_{i=1}^{k} p(\mathbf{z}_i|Z^{i-1}) \tag{10}$$

Given that the above pdfs are Gaussian, and that $E[\mathbf{z}_i] = \mathbf{H}\mathbf{x}_{i|i-1}$, the pdf for each measurement in SLAM is

$$p(\mathbf{z}_i|Z^{i-1}) = N(\tilde{\mathbf{z}}_{i|i-1}; \mathbf{0}, \mathbf{S}_i) = (2\pi)^{-\frac{\dim \mathbf{z}}{2}}|\mathbf{S}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{\mathbf{z}}_{i|i-1}^{\top}\mathbf{S}^{-1}\tilde{\mathbf{z}}_{i|i-1})} \tag{11}$$

That is, the joint pdf of the sequence of measurements Z^k is equal to the product of the marginal pdfs of the corresponding innovations. In practice however, it is more convenient to consider the log likelihood function $\ln \Lambda(\mathbf{x})$. The maximum of $\ln \Lambda(\mathbf{x})$ is at the value of the state $\mathbf{x}$ that most likely gave rise to the observed data Z^k, and is obtained by setting its derivative with respect to $\mathbf{x}$ equal to zero, which gives

$$\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}) = \sum_{i=1}^{k} \mathbf{H}^{\top}\mathbf{S}_i^{-1}\tilde{\mathbf{z}}_{i|i-1} \tag{12}$$

An intuitive interpretation of the maximum of the log-likelihood is that the best estimate for the state $\mathbf{x}$, in the least squares sense, is the one that makes the sum of the entire set of Mahalanobis distances $\sum_{i=1}^{k} \tilde{\mathbf{z}}_{i|i-1}^{\top}\mathbf{S}_i^{-1}\tilde{\mathbf{z}}_{i|i-1}$ as small as possible. A measure that is consistent with the spatial compatibility test described in [21].

The Fisher information matrix, a quantification of the maximum existing information in the observations about the state $\mathbf{x}$, is defined (in [20] and [22]) as the expectation on the dyad of the gradient of $\ln \Lambda(\mathbf{x})$:

$$\mathbf{J} = E[(\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}))(\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}))^{\top}] \tag{13}$$

Taking the expectation on the innovation error $E[\tilde{\mathbf{z}}_{i|i-1}\tilde{\mathbf{z}}_{i|i-1}^{\top}] = \mathbf{S}_i$ in the above formula gives the sum

$$\mathbf{J} = \sum_{i=1}^{k} \mathbf{H}^{\top}(\mathbf{H}\mathbf{P}\mathbf{H}^{\top} + \mathbf{W})^{-1}\mathbf{H} \tag{14}$$

It is easy to verify that in the linear case, this expression for the total Fisher information is only a function of $\mathbf{P}_{r,0|0}$, $\mathbf{V}$, and $\mathbf{W}$. If, on the other hand, the Extended Kalman Filter is used, the Jacobian $\mathbf{H}$ in (14) should be evaluated at the true value of the states $\mathbf{x}_0, \ldots \mathbf{x}_k$. Since these are not available, an approximation is obtained at the estimates

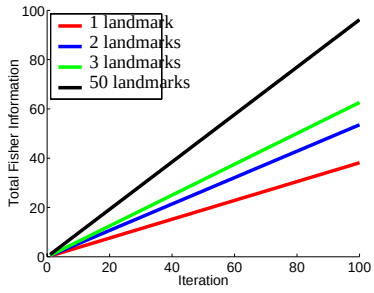


Fig. 3. First entry in the total Fisher information matrix ($\sum\sum\varsigma_{ij}$) for a monobot with variance parameters $\mathbf{P}_{r,0|0} = \mathbf{V} = \mathbf{W} = 1$, and various sizes for the measurement vector.

$\mathbf{x}_{i|i-1}$. The pre and post multiplying $\mathbf{H}$ is, in this context, also known as the *sensitivity matrix* [23].

A necessary condition for the estimator (the Kalman filter) to be consistent in the mean square sense is that there must be an increasing amount of information about the state $\mathbf{x}$ in the measurements. That is, as $k \to \infty$, the Fisher information must also tend to infinity. Fig. 3 shows this for the monobot with constant parameters $\mathbf{P}_{r,0|0} = \mathbf{V} = \mathbf{W} = 1$, and various sizes for the observation vector. Notice how, as the total number of landmarks grows, the total Fisher information also grows, directly relating the number of landmarks to the amount of information available for state estimation in SLAM.

Solving for the k-th sum term in $\mathbf{J}$ for the monobot yields the expression

$$\mathbf{J}_k = \begin{bmatrix} \sum\sum\varsigma_{ij} & -\boldsymbol{\varsigma} \\ -\boldsymbol{\varsigma}^\top & \mathbf{S}_k^{-1} \end{bmatrix} \tag{15}$$

with ς_{ij} the ij-th entry in $\mathbf{S}_k^{-1}$, and $\boldsymbol{\varsigma} = [\sum\varsigma_{1i}, \ldots, \sum\varsigma_{ni}]$.

Citing Bar-Shalom *et al.* [20]: "*A lower bound on the minimum achievable covariance in state estimation is given by the posterior Cramer Rao lower bound*"[3]

$$E[\tilde{\mathbf{x}}_{k+1|k+1}\tilde{\mathbf{x}}_{k+1|k+1}^\top] \geq \mathbf{J}^{-1} \tag{16}$$

Unfortunately, it can be easily shown, at least for the monobot case, that the first row (or column) of $\mathbf{J}$ is equivalent to the sum of the rest of the rows (or columns), producing a singular total Fisher information matrix. In SLAM the Cramer Rao lower bound cannot be evaluated. SLAM is unobservable.

Citing once more Bar-Shalom *et al.*: "*if the Fisher information matrix is not invertible, then the lower bound from (16) will not exist, actually it will have one or more infinite eigenvalues* (one in the case of SLAM), *which means total uncertainty in a subspace of the state space, that is, the information is insufficient for the estimation problem at hand.*"

[3] The matrix inequality $\mathbf{A} \geq \mathbf{B}$ is to be interpreted as $\mathbf{C} = \mathbf{A} - \mathbf{B} \geq 0$. That is, the difference $\mathbf{C}$ of the two matrices is positive semi-definite.

This is a consequence of the form of the Jacobian $\mathbf{H}$, i.e, of the full correlation in SLAM. Zero eigenvalues of $\mathbf{H}^\top \mathbf{S}^{-1}\mathbf{H}$ are an indicator of partial observability, and the corresponding eigenvectors give the unobservable directions in state space.

So for example, for a one-landmark monobot, the innovation covariance is the scalar $s = \sigma_r^2 - 2\rho_{rf}\sigma_r\sigma_f + \sigma_f^2 + \sigma_w^2$, and since $\mathbf{H} = [-1, 1]$, the Fisher information matrix in (14) evaluates to

$$\mathbf{J} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \sum_{i=1}^{k} \frac{1}{s_i} \tag{17}$$

The unobservable direction of the state space is the eigenvector associated to the null eigenvalue of $\mathbf{J}$, we denote it $\mathbf{E}_{\mathrm{Ker}\mathcal{O}}$, since it represents a basis for the Kernel of the observable subspace, and evaluates to

$$\mathbf{E}_{\mathrm{Ker}\mathcal{O}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{18}$$

In Fig. 4 we have plotted the results of using the original fully correlated approach to SLAM for a monobot that starts at location $\mathbf{x}_{r,0|0} = -1m$, and moves along a straight line with a temporal sinusoid trajectory returning to the same point after 100 iterations. Landmarks are located at $\mathbf{x}_{f^{(i)}} = 1m$. A plant noise model proportional to the motion command, and a measurement noise model proportional to the distance from the sensor to the landmark are used. The dotted line indicates 2σ bounds on the state estimate.

The effect of partial observability manifests itself in the dependence on the initial conditions. Note how both the vehicle and landmark mean localization errors do not converge to zero. Their steady state value is subject to the error incurred at the first observation. That is, the filter is unstable. Partial controllability on the other hand, produces a zero Kalman gain for the revision of the landmark estimates. That is, after a few iterations the Kalman filter believes it has a perfectly accurate estimate of the landmark locations, contradictory to the localization error just described. The rate at which the landmark localization Kalman gain approaches zero is dictated by the rate of convergence of the system, i.e., the system's time constant.

A Montecarlo simulation over 100 SLAM runs showed however filter unbiasedness, a property of optimal stochastic state estimation (Kalman filter). That is, the average landmark localization error over the entire set of simulations was still zero, thanks to the independence of the initial landmark measurement errors at each test run. Moreover, the steady state error for the robot and landmark localization is less sensitive to the initial conditions when a large number of landmarks are used. The reason is the same as for the Montecarlo simulation, the observations are independent, and their contribution averages at each iteration in the computation of the localization estimate.

In order to gain full observability we propose to extend the measurement model doing away with the constraint imposed by full correlation. We present two techniques to achieve this. One is to let one landmark serve as a fixed global reference, with its localization uncertainty independent of the vehicle pose. The second proposed technique is the addition of a fixed external sensor, such as a camera, a GPS, or a compass, that can measure all or part of the vehicle location state at all times, independent of the landmark estimates. Both techniques are based essentially on the same principle. Full observabil-

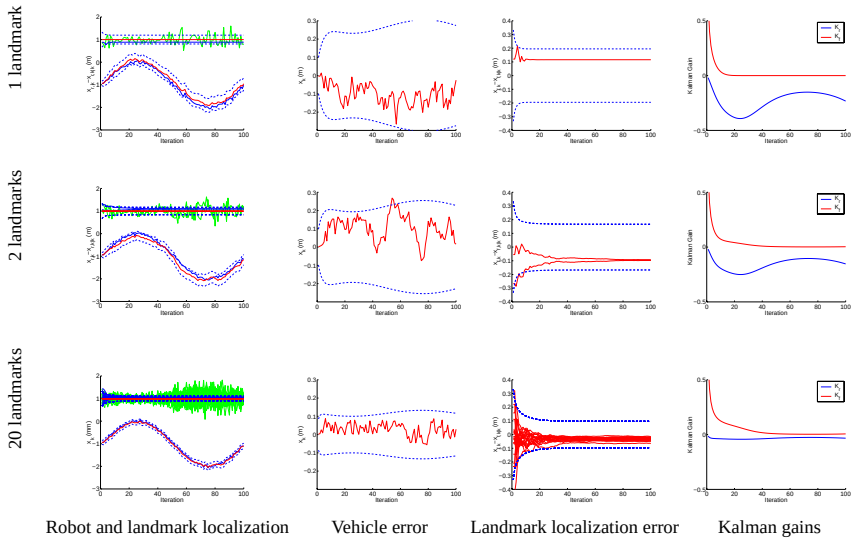


Fig. 4. Full-covariance EKF SLAM for a monobot in a sinusoidal path from $\mathbf{x}_{r,0|0} = -1m$ to $\mathbf{x}_{100} = -1m$ with 100 iterations. The noise corrupted sinusoidal vehicle trajectory is indicated by the darkest curve in the first column of plots. In the same set of figures, and close to it is a lighter curve that shows the vehicle location estimate as computed by the filter, along with a pair of dotted lines indicating 2σ bounds on such estimate. The dark straight lines at the $1m$ level indicate the landmark loction estimates; and the lighter noise corrupted signals are used to represent sensor measurements. Also shown, are a pair of dotted lines for 2σ bounds on the landmark location estimates. The second column of plots shows the vehicle localization error and its corresponding covariance, also on the form of 2σ dotted bounds. The third column shows the same for the landmark estimates. And, the last column contains on dark the vehicle correction Kalman gain, whereas on light, the landmark correction Kalman gain.

ity requires an uncorrelated measurement Jacobian, or equivalently, a full rank Fisher information matrix.

We next present, without loss of generality, the extensions to the monobot SLAM model in order to obtain full observability.

A fixed global reference The plant model is left untouched, i.e., from equation (1),

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{u}_k + \mathbf{v}_k \tag{19}$$

The measurement model takes now the form

$$\begin{bmatrix} z_k^{(0)} \\ \mathbf{z}_k \end{bmatrix} = \begin{bmatrix} -1 & \mathbf{0}_{1\times n} \\ -\mathbf{1}_{n\times 1} & \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} w_k^{(0)} \\ \mathbf{w}_k \end{bmatrix} \tag{20}$$

One of the observed landmarks is to be taken as a global reference at the world origin. No map state is needed for it. The zero-th superscript in the measurement vector

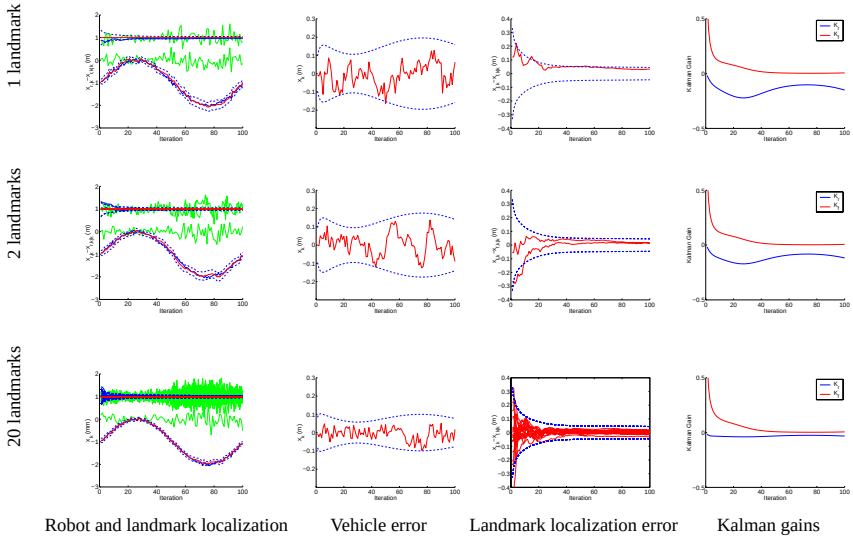


Fig. 5. Full-covariance fully observable SLAM for a monobot in a sinusoidal path from $\mathbf{x}_{r,0|0} = -1m$ to $\mathbf{x}_{100} = -1m$ with 100 iterations. The global reference is observed at the origin.

is used for the consistent indexing of landmarks and observations with respect to the original model. It can be easily shown that the observability matrix for this new model is full rank.

The innovation covariance matrix for the augmented system $\mathbf{S}_{O,k}$ is of size $(n+1) \times (n+1)$, and its inverse can be decomposed in

$$\mathbf{S}_{O,k}^{-1} = \begin{bmatrix} \varsigma_{O,00} & \varsigma_{O,01} \cdots \varsigma_{O,0n} \\ \varsigma_{O,01} & \\ \vdots & \hat{\mathbf{S}}_k^{-1} \\ \varsigma_{O,0n} & \end{bmatrix} \tag{21}$$

with $\varsigma_{O,ij}$ the ij-th entry in $\mathbf{S}_{O,k}^{-1}$, $\boldsymbol{\varsigma}_O = [\sum \varsigma_{O,1i}, \ldots, \sum \varsigma_{O,1i}]$, and $\hat{\mathbf{S}}_k^{-1}$ its submatrix associated to the landmarks that are under estimation (excluding the anchor observation).

The k-th element of the Fisher information matrix sum (14) is now

$$\mathbf{J}_{O,k} = \begin{bmatrix} \sum\sum \varsigma_{O,ij} & -\boldsymbol{\varsigma}_O \\ -\boldsymbol{\varsigma}_O^\top & \hat{\mathbf{S}}_k^{-1} \end{bmatrix} \tag{22}$$

Unlike in (15), this form of the Fisher information matrix is full rank. Moreover, from the properties of positive definite matrices, if $\mathbf{J}_{O,k}$ is positive definite, the entire sum that builds up $\mathbf{J}_O$ is also positive definite.

Fig. 5 shows the results of applying full observability to the same monobot model as the one portrayed in Fig. 4. Note how the steady state (robot pose and landmark locations) is now unbiased with respect to the initial landmark estimates.

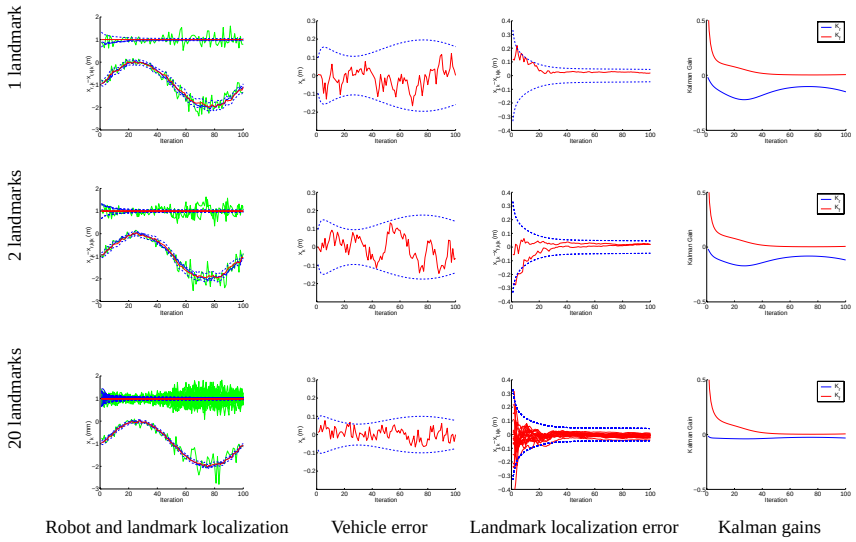


Fig. 6. Full-covariance fully observable SLAM for a monobot in a sinusoidal path from $\mathbf{x}_{r,0|0} = -1m$ to $\mathbf{x}_{100} = -1m$ with 100 iterations. A fixed external sensor is used for the measurement of the vehicle pose.

An external sensor Instead of using one of the landmarks as a global reference, one could also use a fixed sensor to measure the position of the robot. For example, by positioning a camera that observes the vehicle at all times. For such cases, the monobot measurement model may take the form

$$\begin{bmatrix} z_k^{(0)} \\ \mathbf{z}_k \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_{1\times n} \\ -\mathbf{1}_{n\times 1} & \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} w_k^{(0)} \\ \mathbf{w}_k \end{bmatrix} \tag{23}$$

The characteristics of the observability matrix, and the Fisher information matrix, are exactly the same as for the previous case. This new model is once more, fully observable. Fig. 6 shows the results of using an external sensor to measure the vehicle pose. The results are theoretically equivalent to the previous case. The choice of one technique over the other would depend on the availability of such external sensor, and on its measurement noise covariance characteristics.

Planar vehicle The results from the previous section are easily extensible to more complicated vehicle models. For example, the measurement model of a global reference fixed at the origin, for the nonlinear vehicle from Fig. 2 is

$$\mathbf{h}^{(0)} = -\mathbf{R}^\top \mathbf{t} + \mathbf{w}^{(0)} \tag{24}$$

and its corresponding Jacobian is

$$\mathbf{H}_{\mathbf{x}}^{(0)} = [\, -\mathbf{R}^\top \; -\dot{\mathbf{R}}^\top \mathbf{t} \; \mathbf{0}_{2\times 2n} \,] \tag{25}$$

a) Partially Observable SLAM b) Fully Observable SLAM

Fig. 7. EKF SLAM for a planar vehicle.

The case of the external sensor is even simpler, the corresponding equations are

$$\mathbf{h}^{(0)} = \mathbf{t} + \mathbf{w}^{(0)} \tag{26}$$

$$\mathbf{H}_{\mathbf{x}}^{(0)} = [\,\mathbf{I}\ \mathbf{0}_{2\times(2n+1)}\,] \tag{27}$$

In both cases, the symbolic manipulation of (25) and (27) with a commercial algebra package, produce full rank observability matrices. That is, for the planar mobile robot platform used, only one two-dimensional global reference, or the use of a sensor that can measure the xy position of the robot, are sufficient to attain full observability in SLAM. Figure 7a shows a run of the EKF SLAM algorithm over a $300m^2$ area in the second floor of the USC SAL building with data acquired with a laser range finder[4]. The effects of nonlinearities, together with the marginal stability of the filter produce large localization errors at the end of a nearly $90m$ loop. The map obtained with the fully observable SLAM algorithm for the same data set is shown in Figure 7b. In this run, the first observed landmark is used as an anchor. Furthermore, and to guarantee full observability during the entire run of the algorithm, every landmark revised more than 50 times by the filter was removed from the state vector, but its observations were still used to revise the remaining state vector elements.

3 Fusion of Color and Shape for Object Tracking under Varying Illumination

Color represents a visual feature commonly used for object detection and tracking systems, specially in the field of human-computer interaction. For such cases in which the environment is relatively simple, with controlled lighting conditions and an uncluttered background, color can be considered a robust cue. The problem appears when we are dealing with scenes with varying illumination conditions and confusing background. See for example, Fig. 8, with some frames and the corresponding color distributions (in RGB color space), from a motion sequence of a reddish Lambertian surface, in which

[4] Data from the Robotics Data Set Repository [24]. Thanks to Andrew Howards.

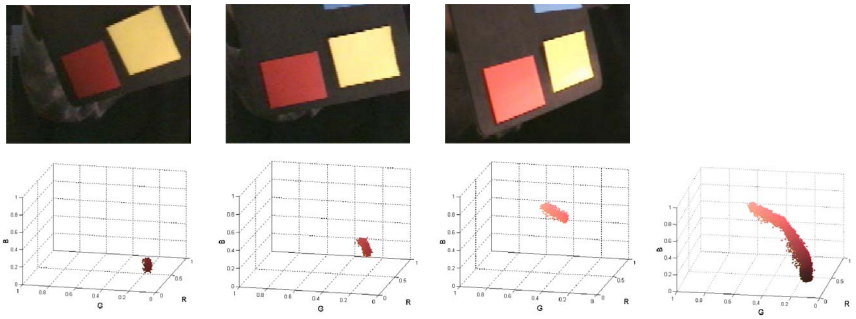


Fig. 8. Example frames of a time-varying color illuminant.

the object of interest revolves around the light source. The path followed by the color distribution for the entire sequence is shown in the last plot.

Thus, an important challenge for any color tracking system to work in real unconstrained environments, is the ability to accommodate variations in the amount of source light reflected from the tracked surface. The choice of different color spaces like HSL, normalized color rgb, or the color space $(B-G, G-R, R+G+B)$, can give some robustness against varying illumination, highlights, interreflections or changes in surface orientation. But none of these transformations is general enough to cope with arbitrary changes in illumination.

Instead of searching for color constancy, other approaches try to adapt the color distribution over time. One such technique is to use Gaussian mixture models to estimate densities of color, and under the assumption that lighting conditions change smoothly over time, the models are recursively adapted. Another option is to parameterize the color distribution as a random vector and to use a second order Markov model to predict the evolution of the corresponding color histogram. These techniques perform much better than the mere change of color space representation, but have the drawback that they do not check for the goodness of the adaptation, which can still lead to failure.

The fusion of several visual modules using different criteria offers more reliability than methods that only use one feature. As an example, systems that track in real-time an individual might model the head of a person by an ellipse and use intensity gradients and color histograms to update the head position over time. In [12], color histograms are fused with stereovision information in order to dynamically adapt the size of the tracked head. These real time applications however are constrained only to the tracking of elliptical shapes.

A new methodology that addresses the problems present in the approaches described above, results in a robust tracking system able to cope with cluttered scenes and varying illumination conditions. The fusion is done using the CONDENSATION algorithm that formulates multiple hypothesis about the estimation of the object's color distribution and validates them taking into account the contour information of the object [25].

3.1 Method Overview

Before entering into a detailed description of the proposed method we give a short glimpse of its main features:

- **Integration of color and shape information:** fusion of both vision modules makes our method appropriate to work in cluttered scenes. If the color distribution of the tracked subject is known, the image can be segmented using color histograms [26], and this information can be used to discriminate many false edges.
- **Ability to adapt shape deformation and varying illumination:** accommodation to varying illuminating conditions is needed to get a good color segmentation of the tracked object. Color segmentation is used to eliminate many false edges from the region of interest, simplifying a final stage of adapting a snake (maintaining affinity) to the contour of the object (assuming that the set of possible shapes of image contours do indeed form affine spaces). We introduce a restriction to the classical snake minimization procedure [27], to obtain affine deformations only. This feature makes our system robust to partial occlusions of the target.
- **Fusion of color and shape in a probabilistic framework:** the CONDENSATION algorithm offers the appropriate framework to integrate both color and contour information, and to perform tracking of the object color distribution in color space, and that of the object contour in image space, both simultaneously. That is, using the predictive filter, multiple estimates of the object color distribution are formulated at each iteration. These estimates are weighted and updated taking into account the object shape, enabling the rejection of objects with similar color but different shape than the target. Finally, the best color distribution is used to segment the image and refine the object's contour.

3.2 Color module

Our color module is highly inspired in Birchfield's real-time head tracking system [28], where the projection of a head in the image plane is modeled by an ellipse. We initialize the process by detecting a human head on an image, using the technique described in our previous work [29]. This method, executed off-line, gives an initial position and scale of the subject head on the image, and lets us construct a model color histogram by filling the buckets of a discretized color space (B-G,G-R,B+G+R), with the pixels inside the ellipse. To cope with situations where the subject turns around, we use a bimodal histogram containing skin and hair data samples.

At run time, when a new image is presented, a head candidate is searched on a local region around the previous position trying to maximize the intersection between the model histogram *M* and the candidate histogram C. The size of the candidate ellipse is given by the previous iteration of the stereo module. $C(i)$ and $M(i)$ represent the number of pixels inside the i-th bucket of the candidate and model histograms respectively, with N the total number of buckets. Swain and Ballard [26] propose the following expression as a measure of histograms intersection:

$$\phi(C) = \frac{\sum_{i=1}^{N} \min\left(C(i), M(i)\right)}{\sum_{i=1}^{N} C(i)} \tag{28}$$

a) rotations b) occlusions c) Other objects

Fig. 9. Face tracking under varying illumination conditions. Robustness under complex situations.

In the search for an ellipse that maximizes the intersection function in an interest region, there will be an overlapping between adjacent ellipses, meaning that adjacent ellipses have common pixels. This redundancy can be exploited since for each new candidate ellipse, its color histogram can be computed from the adjacent ellipse by only subtracting those common pixels of the histogram and adding the new ones. We have adopted this strategy from Birchfield's work, in order to fulfill real-time results.

To obtain a model histogram robust to varying illumination [30], we have updated it over time with the equation:

$$M_k(i) = (1-a)M_{k-1}(i) + (a)C_k(i) \qquad i = 1..N \tag{29}$$

In order to avoid updating with false head candidates, equation (29) is applied only when the measure of the model and candidate intersection is above an empirically determined threshold.

3.3 The Tracking Algorithm

In this section a detailed description of the steps used in the method is presented. For ease of explanation these steps are divided as in the CONDENSATION algorithm (Fig. 10 shows the one dimensional case):

Parameterization and *pdf* of color distribution It has been pointed out that an interesting feature of the presented method is that tracking is performed simultaneously in both color and image spaces. In fact, the element being directly tracked by the filter is the object color distribution $\mathcal{C}$, that at time t is the collection of all image pixel color values I_t that belong to the target, i.e, $\mathcal{C}_t = \{(R_i, G_i, B_i) \mid \ i = 1, \ldots, M_t\}$, where M_t is the number of object points at time t, and $0 \leq R_i, G_i, B_i \leq 1$ (we assume without loss of generality, that the color space is RGB, but it is extensible to any color space). As the set of object points can be arbitrarily high, the state vector $\mathbf{x}_t$ will be a parameterization of $\mathcal{C}_t$ with components (adapted from [30]) $\mathbf{x}_t = \left[\mathbf{m}_t^\top, \boldsymbol{\lambda}_t^\top, \theta_t, \phi_t\right]^\top$, where $\mathbf{m}_t = \left[\bar{R}, \bar{G}, \bar{B}\right]^\top$ is the centroid of $\mathcal{C}_t$, $\boldsymbol{\lambda}_t = [\lambda_1, \lambda_2, \lambda_3]^\top$ are the magnitudes of the principal components of $\mathcal{C}_t$; and θ_t, ϕ_t are the angles centered at $\mathbf{m}_t$ that align the two

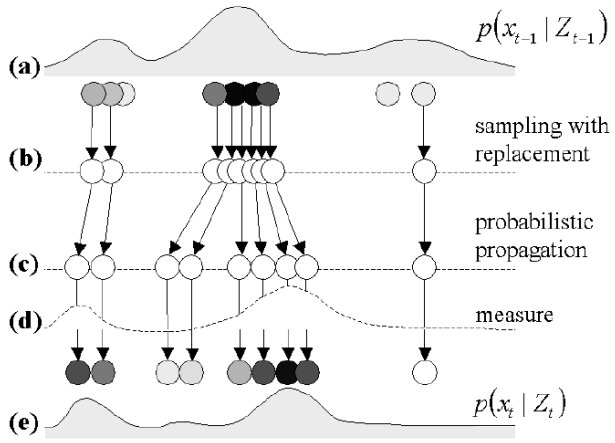


Fig. 10. One iteration of the CONDENSATION algorithm for the one-dimensional case. The weight of each sample is represented by its gray level.

most significant principal components of $\mathcal{C}_t$ with respect to the principal components of $\mathcal{C}_{t-1}$.

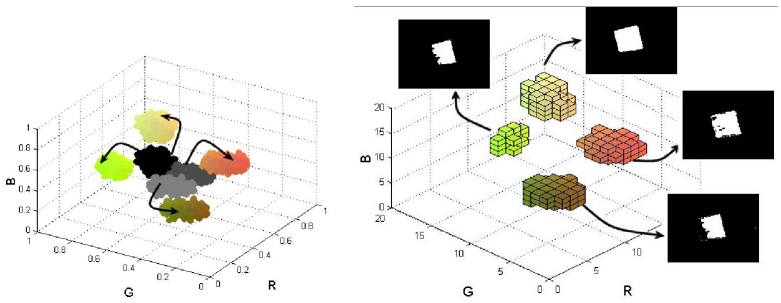
At time t, a set of N samples $\mathbf{s}_{t-1}^{(n)}$ $(n = 1, \ldots, N)$ of the form of $\mathbf{x}$, parameterizing N color distributions $\mathcal{C}_{t-1}^{(n)}$ are available (step (a) from Fig. 10). Each distribution has an associated weight $\pi_{t-1}^{(n)}$. The whole set represents an approximation of the *a posteriori* density function $p\,(b\mathbf{x}_{t-1}|\mathcal{Z}_{t-1})$, where $\mathcal{Z}_{t-1} = \{z_0, \ldots, z_{t-1}\}$ is the history of measurements.

Sampling from $p\,(\mathbf{x_{t-1}}|\mathcal{Z}_{\mathbf{t-1}})$ The next step in the estimation of $p\,(\mathbf{x}_t|\mathcal{Z}_t)$ consists of sampling with replacement N times the set $\left\{\mathbf{s}_{t-1}^{(n)}\right\}$, where each element has probability $\pi_{t-1}^{(n)}$ of being chosen (step (b) from Fig. 10). This, will give us a new set $\left\{\mathbf{s'}_t^{(n)}\right\}$ of color distribution parameterizations. Those distributions having higher weights may be chosen several times, so the new set can have identical copies of elements. On the other hand, those distributions having lower weights may not be chosen (see Fig. 11a).

Probabilistic propagation of samples Each sample $\mathbf{s'}_t^{(n)}$ of the set is propagated (see Fig. 10c and Fig. 11a) according to the following dynamic model:

$$\mathbf{s}_t^{(n)} = A\mathbf{s'}_t^{(n)} + B\mathbf{w}_t^{(n)}$$

where A is the deterministic part, assigned as a first order model describing the movement of an object with constant velocity. $B\mathbf{w}_t^{(n)}$ is the stochastic component, with $\mathbf{w}_t^{(n)}$ a vector of standard normal random variables with unit standard deviation, and $BB^\top$



(a) (b)

Fig. 11. (a) Sampling and probabilistic propagation from color distributions $\mathcal{C}_t^{(n)}$. Observe that the sample having the highest weight has been chosen two times, while another sample with lower weight has not been chosen. (b) Construction of the histograms $\mathcal{H}_t^{(n)}$ and results of the corresponding segmentations $\mathcal{S}_t^{(n)}$.

is the process noise covariance. The parameters A and B are estimated a priori from a training sequence.

Each predicted sample $\mathbf{s}_t^{(i)}$ represents the set of parameters defining the rigid transformations that will be used to warp the color distribution $\mathcal{C}'^{(i)}_t$ associated with the sample $\mathbf{s}'^{(i)}_t$, in order to obtain the new estimated distribution $\mathcal{C}_t^{(i)}$ (with parameters $\mathbf{s}_t^{(i)}$).

Measure and weight In this step, each element $\mathbf{s}_t^{(n)}$ has to be weighted according to some measured features, and is precisely at this point where we integrate the structural information of the object's contour. From the propagated color distributions $\mathcal{C}_t^{(n)}$, we construct the color histograms $\mathcal{H}_t^{(n)}$ with $\mathcal{R} \cdot \mathcal{G} \cdot \mathcal{B}$ bins:

$$\mathcal{H}_t^{(n)}(r,g,b) = \#\left\{ (R,G,B) \in \mathcal{C}_t^{(n)} \mid \frac{r-1}{\mathcal{R}} < R \leq \frac{r}{\mathcal{R}}, \frac{g-1}{\mathcal{G}} < G \leq \frac{g}{\mathcal{G}}, \frac{b-1}{\mathcal{B}} < B \leq \frac{b}{\mathcal{B}} \right\}$$

and where $r = [1,\ldots,\mathcal{R}]$, $g = [1,\ldots,\mathcal{G}]$, $b = [1,\ldots,\mathcal{B}]$, with $\{r,g,b,\mathcal{R},\mathcal{G},\mathcal{B}\} \in \mathbb{N}$. This histogram is used to generate a segmentation $S_t^{(n)}$ from the entire image I_t. That is, given a pixel $I_t(u,v)$ with color value (R,G,B) the corresponding value of the segmented image $S_t^{(n)}(u,v)$ will be assigned a value 1 if $H_t^{(n)}(r,g,b) > 0$, where $r = \lfloor R \cdot \mathcal{R} \rfloor$, $g = \lfloor G \cdot \mathcal{G} \rfloor$ and $b = \lfloor B \cdot \mathcal{B} \rfloor$ (Fig. 11b).

The goal is to assign higher weights to the samples $\mathbf{s}_t^{(n)}$ generating "better" segmentations of the tracked object. To this end, simple morphological operations are performed on $S_t^{(n)}$ to extract a blob corresponding to the segmented object (Fig. 12a). After adjusting a snake along the contour of this blob, the weight assigned to $\mathbf{s}_t^{(n)}$ is computed according to the function:

$$\pi_t^{(n)} = e^{-\frac{\rho^2}{2\sigma^2}}$$

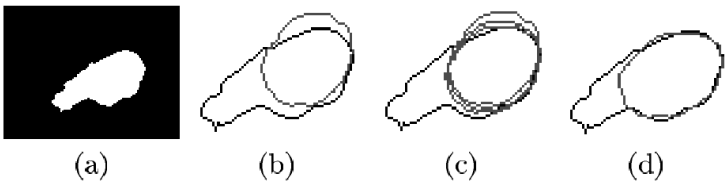


Fig. 12. (a) Segmented image $S_t^{(n)}$ of a snail after postprocessing operations (b) The ellipse represents the initial snake (r_{t-1}). The other curve is the boundary ($\nabla S_t^{(n)}$) of the segmented image. (c) Intermediate steps of the affine snake fitting. (d) Final result of the snake fitting ($r_t^{(n)}$). Observe that if the deformations were not affine, the snake may have erroneously evolved to encompass the neck of the snail.

and $\rho = \mu_1(1 - \Phi_{affine}) + \mu_2(1 - \Phi_{area}) + \mu_3(1 - \Phi_{quality})$.

Functions Φ_{affine}, Φ_{area} and $\Phi_{quality}$ return a value in the range $[0, 1]$ and represent a measure of the following features:

– **Affine similarity**: let $\nabla S_t^{(n)}$ be a binary image of the edges of $S_t^{(n)}$, and r_{t-1} a collection of image points along the snake adjusted to the contour of the object in the iteration $t - 1$. r_{t-1} is used as initialization of an affine snake $r_t^{(n)}$ that is adjusted to $\nabla S_t^{(n)}$. Φ_{affine}, measures the similarity between $r_t^{(n)} = (u_t^{i,(n)}, v_t^{i,(n)})$ $(i = 1, ..., N_r)$, and $\nabla S_t^{(n)}$:

$$\Phi_{affine} = \frac{1}{N_r} \sum_{i=1}^{N_r} \nabla S_t^{(n)} \left(u_t^{i,(n)}, v_t^{i,(n)} \right)$$

– **Congruent value of area**: another factor to take into account when evaluating the goodness of the segmentation $S_t^{(n)}$ is how close is the area $Area_t^{(n)}$ of the snake $r_t^{(n)}$ to the predicted area $\widetilde{Area}_t = Area_{t-1} + \mu(Area_{t-1} - Area_{t-2})$, where $Area_{t-i}$ is the area of the refined snake at iteration $t - i$. This is considered in the function:

$$\Phi_{area} = |\widetilde{Area}_t - Area_t^{(n)}| / \max \left\{ \widetilde{Area}_t, Area_t^{(n)} \right\}$$

– **Quality of the segmentation**: the function $\Phi_{quality}$ is introduced to penalize those segmentations of "low" quality that present some holes into the area of the segmented object. $\Phi_{quality}$ is a linear function of the Euler number of the processed $\mathcal{S}_t^{(n)}$.

Finally, the set of N weights $\pi_t^{(n)}$ associated to each of the samples $\mathbf{s}_t^{(n)}$, represents an approximation to the *a posteriori* density function $p(\mathbf{x}_t | \mathcal{Z}_t)$.

Contour updating The last step of our algorithm, consists in refining the fitting of the object boundary, in order to obtain r_t. This is done by taking the contour of the

segmented image corresponding to the best sample ($\nabla S_t^{(i)} \mid \pi_t^{(i)} \geq \pi_t^{(j)} \ \forall j \neq i, 1 \leq j \leq N$), and instead of adjusting the snake r_{t-1} to $\nabla S_t^{(i)}$, it is adjusted to $\nabla I_t^* = \nabla I_t \cdot dil(\nabla S_t^{(i)})$, where ∇I_t is the gradient of I_t, and the function $dil(\cdot)$, refers to a morphological dilate operation. ∇I_t^* is in fact the original edge image, from which all the clutter and disturbing edges have been eliminated.

3.4 Experimental Results

In this Section four sets of sequence results are presented (summarized in Fig. 13) to illustrate the robustness of our system under different conditions. As the algorithm has been implemented in an interpretative language (MATLAB), speed results will not be analyzed. Attention will be focused on the effectiveness of the method. In the first experiment we show how our system is able to accommodate color by applying it over a synthetic sequence of circles moving around and changing randomly its color. In the upper left image of Fig. 13 the path of the color distributions for the tracked circle is shown. The second experiment (tracking of a colored rectangle) corresponds to the sequence introduced in Fig. 8. It has to be pointed out that in the previous experiment we used the RGB color space, but in the present and subsequent experiments the color space used was the $(B-G, G-R, R+G+B)$ in order to provide robustness to specular highlights. The last two experiments, correspond to outdoor scenes, where although the change in illumination conditions is limited, they are useful to show that our method works with non-uniform shapes (third experiment of a beatle tracking), and in cluttered scenarios (fourth experiment of a snail tracking).

References

1. Andrade-Cetto, J., Sanfeliu, A.: Concurrent map building and localization on indoor dynamic environments. Int. J. Pattern Recogn. **16** (2002) 361–374
2. Andrade-Cetto, J.: Environment Learning for Indoor Mobile Robots. PhD thesis, UPC, Barcelona (2003)
3. Andrade-Cetto, J., Sanfeliu, A.: The effects of partial observability in SLAM. In: Proc. IEEE Int. Conf. Robot. Automat., New Orleans (2004) 397–402
4. Vidal-Calleja, T., Andrade-Cetto, J., Sanfeliu, A.: Estimator stability analysis in SLAM. Technical Report IRI-DT-04-01, IRI, UPC (2004) To be presented at *5th Symposium on Intelligent Autonomous Vehicles*.
5. Vidal-Calleja, T., Andrade-Cetto, J., Sanfeliu, A.: Conditions for suboptimal filter stability in SLAM. To be presented at *IEEE International Conference on Intelligent Robots and Systems* (2004)
6. Andrade-Cetto, J., Sanfeliu, A.: Concurrent map building and localization with landmark validation. In: Proc. 16th IAPR Int. Conf. Pattern Recog. Volume 2., Quebec, IEEE Comp. Soc. (2002) 693–696
7. Andrade-Cetto, J., Sanfeliu, A.: Concurrent map building and localization with temporal landmark validation. In: Proc. IEEE Int. Conf. Robot. Automat., Taipei (2003) 1576–1581
8. Grau, A., Climent, J., Serratosa, F., Sanfeliu, A.: Texprint: A new algorithm to discriminate textures structurally. In Caelli, T., Amin, A., Duin, R.P.W., Kamel, M., de Ridder, D., eds.: Proc. IAPR Int. Workshop Syntactical Structural Pattern Recog. Volume 2396 of Lect. Notes Comput. Sci., Windsor, Springer-Verlag (2002) 368–377

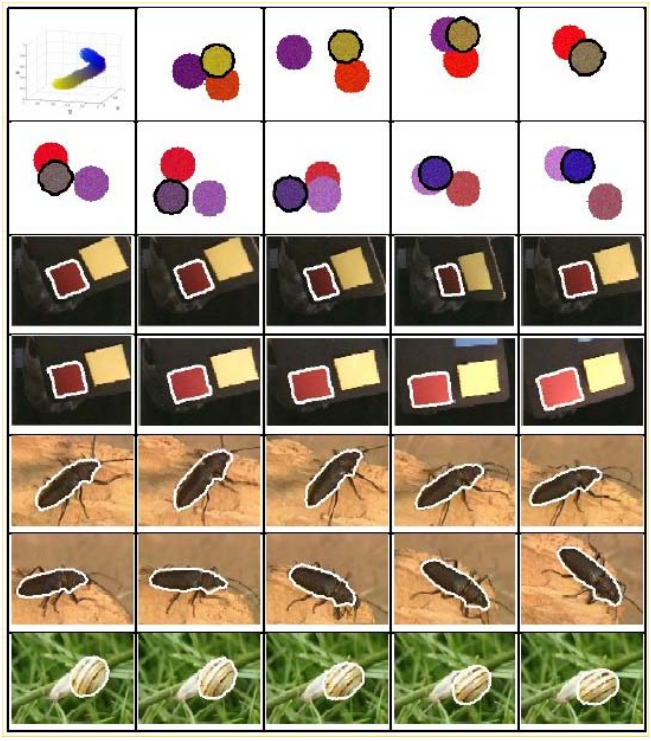


Fig. 13. Results of the 4 experiments.

9. Sanfeliu, A., Alquézar, R., Andrade, J., Climent, J., Serratosa, F., Vergés, J.: Graph-based representations and techniques for image processing and image analysis. Pattern Recogn. **35** (2002) 639–650
10. Vergés-Llahí, J., Tarrida, A., Sanfeliu, A.: New approaches to colour histogram adaptation in face tracking tasks. In: Proc. 16th IAPR Int. Conf. Pattern Recog., Quebec, IEEE Comp. Soc. (2002) 381–384
11. Vergés-Llahí, J., Sanfeliu, A.: Colour constancy algorithm based on the minimization of the distance between colour histograms. In Perales, F.J., Campilho, A.J.C., de la Blanca, N.P., Sanfeliu, A., eds.: Proc. 1st Iberian Conf. Pattern Recog. Image Anal. Volume 2652 of Lect. Notes Comput. Sci., Puerto de Andratx, Springer-Verlag (2003) 1066–1073
12. Moreno, F., Tarrida, A., Andrade-Cetto, J., Sanfeliu, A.: 3d real-time head tracking fusing color histograms and stereovision. In: Proc. 16th IAPR Int. Conf. Pattern Recog. Volume 1., Quebec, IEEE Comp. Soc. (2002) 368–371
13. Moreno, F., Andrade-Cetto, J., Sanfeliu, A.: Fusion of color and shape for object tracking under varying illumination. In Perales, F.J., Campilho, A.J.C., de la Blanca, N.P., Sanfeliu, A., eds.: Proc. 1st Iberian Conf. Pattern Recog. Image Anal. Volume 2652 of Lect. Notes Comput. Sci., Puerto de Andratx, Springer-Verlag (2003) 580–588
14. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Can. Cartographer **10** (1973) 112–122
15. Hershberger, J., Snoeyink, J.: Speeding up the douglas-peucker line-simplification algorithm. In: Proc. 5thInt. Sym. Spatial Data Handling, Charleston (1992) 134–143

16. Checa, A., Andrade, J., Sanfeliu, A.: Construcció de mapes per robots móbils equipats amb sensors làser de profunditat. Technical Report IRI-DT-03-01, IRI, UPC (2003)
17. Smith, R.C., Cheeseman, P.: On the representation and estimation of spatial uncertainty. Int. J. Robot. Res. **5** (1986) 56–68
18. Dissanayake, M.W.M.G., Newman, P., Clark, S., Durrant-Whyte, H.F., Csorba, M.: A solution to the simultaneous localization and map building (SLAM) problem. IEEE Trans. Robot. Automat. **17** (2001) 229–241
19. Gibbens, P.W., Dissanayake, G.M.W.M., Durrant-Whyte, H.F.: A closed form solution to the single degree of freedom simultaneous localisation and map building (SLAM) problem. In: Proc. IEEE Int. Conf. Decision Control, Sydney (2000) 408–415
20. Bar-Shalom, Y., Li, X.R., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation. John Wiley & Sons, New York (2001)
21. Neira, J., Tardós, J.D.: Data association in stochastic mapping using the joint compatibility test. IEEE Trans. Robot. Automat. **17** (2001) 890–897
22. Bar-Shalom, Y., Fortmann, T.H.: Tracking and Data Association. Volume 179 of Mathematics in Science and Engineering. Academic Press, Boston (1988)
23. Dowski, E.R.: An information theory approach to incoherent information processing. In: Proc. 5th Topical Meeting on Signal Recovery and Synthesis. OSA Technical Digest (1995) 106–108
24. Howard, A., Roy, N.: The robotics data set repository (Radish) (2003)
25. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. Int. J. Comput. Vision **29** (1998) 5–28
26. Swain, M., Ballard, D.: Color indexing. Int. J. Comput. Vision **7** (1991) 11–32
27. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Comput. Vision **1** (1987) 321–331
28. Birchfield, S.: Elliptical head tracking using intensity gradients and color histograms. In: Proc. 12th IEEE Conf. Comput. Vision Pattern Recog., Santa Barbara (1998) 232–237
29. Moreno, F., Andrade-Cetto, J., Sanfeliu, A.: Localization of human faces fusing color segmentation and depth from stereo. In: Proc. IEEE Int. Conf. Emerging Tech. Factory Autom., Antibes (2001) 527–536
30. Sigal, L., Sclaroff, S., Athitsos, V.: Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. In: Proc. 14th IEEE Conf. Comput. Vision Pattern Recog. Volume 2., Head Island (2000) 152–159