An information theoretic rule for sample size adaptation in particle filtering

Oswald Lanz*
Bruno Kessler Foundation - irst
Via Sommarive 18, 38050 Trento, Italy

Abstract

To become robust, a tracking algorithm must be able to support uncertainty and ambiguity often inherently present in the data in form of occlusion and clutter. This comes usually at the price of more demanding computations. Sampling methods, such as the popular particle filter, accommodate this capability and provide a means of controlling the computational trade-off by adapting their resolution. This paper presents a method for adapting resolution on-the-fly to current demands. The key idea is to select the number of samples necessary to populate the high probability regions with a predefined density. The scheme then allocates more particles when uncertainty is high while saving resources otherwise. The resulting tracker propagates compact while consistent representations and enables for reliable real time operation otherwise compromised.

1. Introduction

Visual inference in unconstrained scenes is always affected by uncertainty and ambiguity. This is of particular concern when tracking multiple independent bodies: uncertainty and ambiguity may derive from inaccurate interpretation of images, but can also be intrinsic in the measurement process, e.g. when occlusions exist, or in the monitored scene itself, e.g. when different targets appear similar or clutter is present in the background. Even when the interpretation is carried out using a physically based model of the visual observation process [6, 10], information is lost during occlusion and estimations become uncertain. Another critical example is the robot self localization task [5, 4], where redundancy in the visual appearance of the scene is unavoidably reflected in ambiguous and uncertain localization estimates. Bayesian methods allow to account for this in a principled way by maintaining estimates in form of distributions rather than by a single, deterministic output. Propagating such distributions over time then allows for a consistent support of uncertainty and ambiguity, thus accounting in a natural way for measurement noise and multiple hypotheses.

When it comes to implementation, however, a computational problem is faced. Visual likelihoods arising in such scenes are complex and highly non-linear, and so are the resulting distributions. This rules out the use of parametric representations, which are on the basis of the efficient Kalman filter. More suitable are then sample based representations, as adopted in the popular particle filter. The drawback is that such methods usually need a large number of samples to work reliably and are therefore computationally demanding. The key to efficiency is then to continuously adapt the resolution of such estimates to the data at hand to obtain compact while still consistent representations. The contribution of this paper is in this direction and proposes an information-theoretic rule which makes a particle filter based tracker adaptive to its own estimation uncertainty.

This paper is organized as follows. After reviewing the basics of particle filtering and a brief discussion of related work, in Sec. 2 we present the sample size selection scheme, which is the core contribution of this paper. Sec. 3 shows how this rule can be applied to particle filtering to implement sample size adaptation. Sec. 4 presents obtained results on synthetic and real data. Sec. 5 has the conclusions.

1.1. Background

In a Bayesian state estimation setting, the aim is to recursively estimate the posterior distribution $p(x_t|z_{1:t})$ of a vectorial representation x of the state of a monitored environment conditioned on a sequence of observations $\mathbf{z}_{1:t}$. This is done in two steps, by first propagating the posterior obtained at the previous time $p(x_{t-1}|z_{1:t-1})$ according to a model $p(x_t|x_{t-1})$ of expected evolution and then updating it with the information contained in the new observation according to a model $p(z_t|x_t)$ of the observation process:

$$p(x_t|z_{1:t}) \propto p(z_t|x_t) \int p(x_t|x_{t-1}) p(x_{t-1}|z_{1:t-1})\, dx_{t-1}. \quad (1)$$

*The author has been partly funded by the European Union under project CHIL (Computers in the Human Interaction Loop, IP 506909).

When the observation likelihood $p(z_t|x_t)$ is a complex function of x_t, as happens in most vision based state estimation problems, the posterior $p(x_t|z_{1:t})$ cannot be expressed in closed form. One has then to resort to approximations such as the extended Kalman filter, mixture filter [7], PHD filter [14], grid filter [11], MCMC methods [8], or particle filter [1]. The last one has earned great popularity in the last decade, mostly because it can accommodate generic dynamical and likelihood models and builds upon universal representations while remaining simple in implementation.

The idea underlying particle filters is to maintain a compressed representation of the posterior by means of a set of representative sample states, the particles. In perfect Monte Carlo sampling, these samples are chosen independently and identically distributed (i.i.d.) according to the density $p(x)$ that they should represent. If $p(x)$ is difficult to sample, one can sample from some other feasible importance density $g(x)$ while correcting the introduced bias by sample weighting. A set of weighted particles $\{\langle x_i, \pi_i \rangle\}$ with $x_i \sim g(x)$ approximates then an arbitrary density p according to

$$\Pr(\mathcal{A}) = \int_{\mathcal{A}} p(x)\, dx \approx \sum_i \pi_i\, \delta_{\mathcal{A}}(x_i), \quad \pi_i = \frac{p(x_i)}{g(x_i)} \quad (2)$$

with $\delta_{\mathcal{A}}$ denoting the characteristic function of an arbitrary set $\mathcal{A}$. Given a weighted particle representation $\{\langle \bar{x}_i, \bar{\pi}_i \rangle\}$ for the posterior at time $t-1$, Bayes recursion (Eq. 1) becomes

$$p(x_t|z_{1:t}) \approx\ p(z_t|x_t) \sum_i \bar{\pi}_i\, p(x_t|\bar{x}_i). \quad (3)$$

A common choice for the importance density is the mixture density derived from the dynamical model

$$g(x_t) = \sum_i \bar{\pi}_i\, p(x_t|\bar{x}_i). \quad (4)$$

At each iteration t, a new set of representative particles $\{x_i\}$ is i.i.d.-sampled from $g(x_t)$. Then, observation z_t is analyzed for new importance weights $\pi_i = p(z_t|x_i)\bar{\pi}_i$. To focus on likely trajectories, particles are periodically resampled according to their weights [3, 1].

1.2. Related work

Adaptation of sample size in particle filters has been a studied topic in the robotics community. In the likelihood based adaptation scheme [9] new particles are generated until the sum of their likelihoods exceeds a predefined threshold. This method has the advantage of introducing no overhead in the filter iteration, but it may significantly underestimate the required sample size when ambiguity is present in form of multiple modes in the density. A rigorous approach, termed KLD sampling, is presented in [4]. The underlying idea is to bound the Kullback-Leibler (KL) divergence between the true distribution and its incrementally sampled representation. Under the assumption that the true distribution is well modeled by a piecewise constant, multinomial distribution defined over a regular grid, it is possible to derive a closed form expression for the number of samples needed to satisfy a given KL divergence bound. Although this assumption may in theory limit its accuracy, this method shows excellent performance in practice. Recent work [13] extends KLD sampling to account for the inefficiency when sampling from an importance density that has significant offset from the true distribution. A correction factor is found by arguing that Monte Carlo integration and importance sampling estimators should have comparable variance when computing the distribution mean.

2. An information theoretic criterion for sample size determination

The performance of a tracking algorithm varies with many factors. Among them, observation frame rate, background clutter and illumination change do have an important impact even if only a single target is tracked. This problem becomes even more significant when tracking multiple bodies which may partially or even completely occlude each other. It is therefore essential in this context that a system be adaptive to its own estimation uncertainty.

2.1. Uncertainty, entropy and bulkyness of the typical set

Any estimated particle set is essentially a sample based representation of a probability density p. This allows to define uncertainty in terms of entropy

$$\mathcal{H}(p) = -\int p(x) \ln p(x)\, dx. \quad (5)$$

It has many different interpretations, ranging from minimum compression code length to a measure of disorder in thermodynamics. Here entropy will be read as a measure of uncertainty of a continuous density, a quantity that will finally guide us in selecting the appropriate size of particle sets. What we are looking for is a rule that states how many independently and identically distributed samples are needed to populate that portion of space, say A_p, which contains most of the probability mass of p. In information theory A_p is known as the *typical set* of p. Fig. 1 shows the plots of several equi-probable sets with different area and the typical set with the smallest possible one. The following statement is a consequence of the asymptotic equipartition property (AEP) theorem [2] and allows to measure how *bulky* the typical set is:

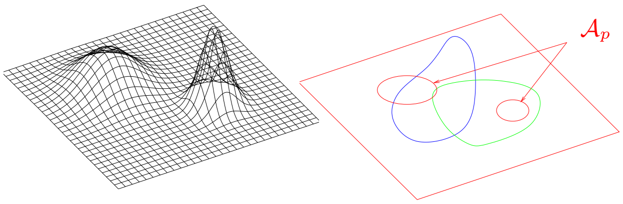

Figure 1. A 2D density and several state space regions with different area but equal mass. Among them, the typical set $\mathcal{A}_p$ is the smallest possible one.

For a given density p and n large, the volume of the smallest n–sized sample set that contains most of the probability is approximately $e^{n\mathcal{H}(p)}$.

2.2. AEP based sample size determination

As a corollary of this theorem it can be proven that, among all n–sized sample sets, the one with highest probability and smallest volume is given by $\mathcal{A}_p \times \ldots \times \mathcal{A}_p$ (n times). This follows from the fact that the samples are supposed to be independently drawn from p and thus their joint space can be factorized. This allows to conclude that $\mathcal{A}_p$, the slice containing most of the probability mass of p, has volume approximately equal to $e^{\mathcal{H}(p)}$. With the purpose of populating the typical set $\mathcal{A}_p$ with a given sample density ρ, the following sampling rule can be enunciated.

AEP based sample size determination: *The number of i.i.d. samples needed to properly represent a density p with resolution ρ is*

$$n_\rho(p) = \rho \,\mathrm{Vol}(\mathcal{A}_p) \approx \rho \, e^{\mathcal{H}(p)}. \tag{6}$$

This rule states that the number of samples required is equivalent to the number needed to uniformly populate the characteristic counterpart of p, i.e. the uniform density with support $\mathcal{A}_p$. While a rigorous enunciate assessing convergence properties is missing in this rule, it inherits its soundness from the AEP theorem, and it has the advantage of being simple and available as an analytical function of $\mathcal{H}(p)$, the formal expression of uncertainty.

3. AEP based sample size adaptation in particle filtering

Unfortunately, the entropy of generic densities does not have analytic form. This holds also for simple mixtures like the temporal prior in Eq. 4 deriving from a Gaussian dynamical model. Instead of referring to demanding numerical methods, an approximation can be computed from the sample set itself, thus in a Monte Carlo manner. This way, entropy can be estimated also when p itself is not available in closed form but can be evaluated point-wise. This is the case for posteriors arising from complex likelihoods typical in appearance based tracking and localization.

3.1. Entropy of a particle set

The implementation of an adaptive propagation scheme based on a sampling rule built on top of Eq. 6 requires to estimate the entropy of the posterior while sampling its weighted particle representation. Thus the question arises how this can be done incrementally, from its sample representations $\{\langle \pi_i, x_i\rangle \,|\, i = 1, \ldots, \mathcal{I}\}$ with increasing resolution $\mathcal{I}$. Computing $\mathcal{H}$ by applying the formula for approximating the expectation of a generic function $f(x)$

$$\int f(x)p(x)\,dx = \mathcal{E}_p\{f(x)\} \approx \sum \pi_i f(x) \tag{7}$$

on p in its discrete (particle) form $\sum \pi_i \delta(x_i)$ can perform poorly

$$\int p(x)\ln p(x)\,dx = \mathcal{E}_p\{\ln p(x)\} \not\approx \sum \pi_i \ln \pi_i. \tag{8}$$

To show this, let us consider the n–sized particle set obtained after applying weighted resampling. In this case all the particles have the same weight $1/n$. Hence, any such particle set (with the same cardinality) would have the same estimated entropy, namely $-\sum 1/n \,\ln(1/n) = \ln n$, no matter if they are all different instances of the same state (a perfect estimate) or if they are spread out over a large area of the state space (a uniform, uninformative belief). The problem is that the above formula does not preserve the spatial relationship between the samples. More formally, $\pi_i f(x_i)$ approximates the contribution of f due to the infinitesimal volume $p(x_i)dx$ rather than the one due to the density value $p(x_i)$ in x_i only (the spatial component dx is lost). This aspect has not always been considered in the literature [12].

A more reliable value is computed by applying Eq. 7 to log-values obtained by means of kernel density estimation

$$\mathcal{H}(p) \approx \sum_i \pi_i \ln \mathcal{E}_p\{K(x_i|\,:)\} \approx \sum_i \pi_i \ln \sum_j \pi_j K(x_i|x_j) \tag{9}$$

where $K(:\,|\,:)$ denotes a suitable radial kernel, e.g. a Gaussian one. The choice of the kernel bandwidth depends on the application at hand, and in particular on requested accuracy and likelihood sharpness. For a 3D people tracking application a suitable value is found to be a portion of the body width, e.g. 10cm. Alternatively, a nearest neighbor technique could be used.

Algorithm 1 Adaptive particle filter iteration

input: previous belief $\{\langle\bar{\pi}_i, \bar{x}_i\rangle\}$, entropy update period Δn, range of admissible particle set sizes $[n_{\text{min}} : n_{\text{max}}]$

burn-in:
sample n_{min} times from prior mixture $\sum \bar{\pi}_i p(x|\bar{x}_i)$
compute n_{min} new likelihoods $\pi_i = p(z_t|x_i)$
set n_{curr} to n_{min} and n_{sat} to n_{max}

iterate:
while ($n_{\text{curr}} < \min\{n_{\text{sat}}, n_{\text{max}}\}$) **do**
sample Δn times from prior mixture $\sum \pi_i p(x|x_i)$
compute Δn new likelihoods $\pi_i = p(z_t|x_i)$
set n_{curr} to $n_{\text{curr}} + \Delta n$
estimate entropy (Eq. 9) and new sample size (Eq. 6)
set n_{sat} to new sample size
correct n_{sat} for inefficiency (e.g. see [13])

output: new belief $\{\langle\pi_i, x_i\rangle\}$

3.2. An adaptive sampling algorithm

An adaptive particle filter implementing the proposed sample sized adaptation scheme is outlined in Alg. 1. After a minimum number of particles have been drawn from the temporal prior (Eq. 4) and their likelihoods evaluated, further samples are generated and scored only until the set becomes saturated according to Eq. 6.

The complexity of evaluating Eq. 9 is quadratic in the number of effective particles (i.e. the number of different particles, without counting multiple instances of the same one). If this overhead is critical, the entropy can be evaluated a priori, for the previously resampled posterior, for which a single state is usually instantiated several times. This value can then be properly amplified according to elapsed time. Motivated by continuity, further savings in early entropy evaluation can be achieved by setting the burn-in size to a value slightly lower than the resolution of the previous particle set. If likelihoods cannot be evaluated while sampling but need to be computed afterwards, the method can still be used to determine the size required to consistently represent the temporal prior. In such case the entropy (Eq. 9) may be computed without deferring to kernel density estimation, by assigning particle densities directly from the mixture in Eq. 4.

The resolution of posterior representation can be controlled by the choice of sampling parameter ρ, as it imposes a more or less dense population of the typical set. Parameter ρ can be adapted according to task-specific demands that may vary in space and time. As an example, in a robot localization task an accurate position estimate might be required at some places of the environment (high ρ near obstacles), while in other regions it could be sufficient to simply keep track without concerns about resolution (low ρ).

4. Experiments

4.1. Quantitative evaluation

To demonstrate the soundness of the approach we analyze how closely the consistently sized particle set approximates the true distribution on three simple 1D examples. To do so we first apply Algorithm 1 to sample the particle set, with prior mixture and likelihood function as specified for each example below. Then we numerically evaluate KL divergence between target density and the continuous density reconstructed from the particle set by kernel density estimation. This comparison is consistent with Eq. 9, which uses the same continuous representation to estimate entropy, and thus sample size. Numerical integration step is 0.01, kernels are Gaussian with standard deviation 0.1, burn-in sample size is 50. Divergence values reported are computed as an average over 50 runs of the algorithm.

For the first experiment we consider Gaussian prior and uniform likelihood. Fig. 2(a) shows the plot of KL divergence between target and reconstructed density as a function of the entropy of the target distribution. For a Gaussian with standard deviation σ the entropy is available in analytic form and reads $\mathcal{H}(p) = \ln(\sigma\sqrt{2\pi e})$. When using constant number of samples ($n = 100, 200, 300, 400$) the accuracy breaks down after a certain amount of uncertainty has been reached. This behavior is as expected, since the extent of the region with high probability increases and the limited number of samples are not able to populate it adequately. The adaptive choice of the sample size keeps the reconstruction error at a low, constant value. Even more: by considering Fig. 2(b) one can conclude that the number of samples drawn is optimal. Indeed, the entropy values after which the reconstruction accuracy breaks down (2(a)) coincide with the entropy values for which the AEP rule suggests precisely that constant amount of particles (2(b)). The graph shows also that the algorithm is able to accurately estimate the entropy of the target distribution since the sampled graph follows tightly the analytic curve. The behavior at low entropy rates (2(a)), increasing accuracy) can be explained by considering the smoothing effect introduced by kernel density estimation which becomes significant for narrow priors.

In the second experiment (Fig. 3) target distribution p is chosen as a mixture of Gaussian and uniform density, $p(x) = 0.5\mathcal{N}(x|0, \sigma) + 0.5\mathcal{U}(x|3 - \sigma, 2\sigma)$. By increasing σ within a range and numerically evaluating entropy, the same KL divergence plot is produced. It confirms robustness to large uncertainty in the target distribution when using AEP rule. The behavior around $\mathcal{H}(p) = 1.45$ can be

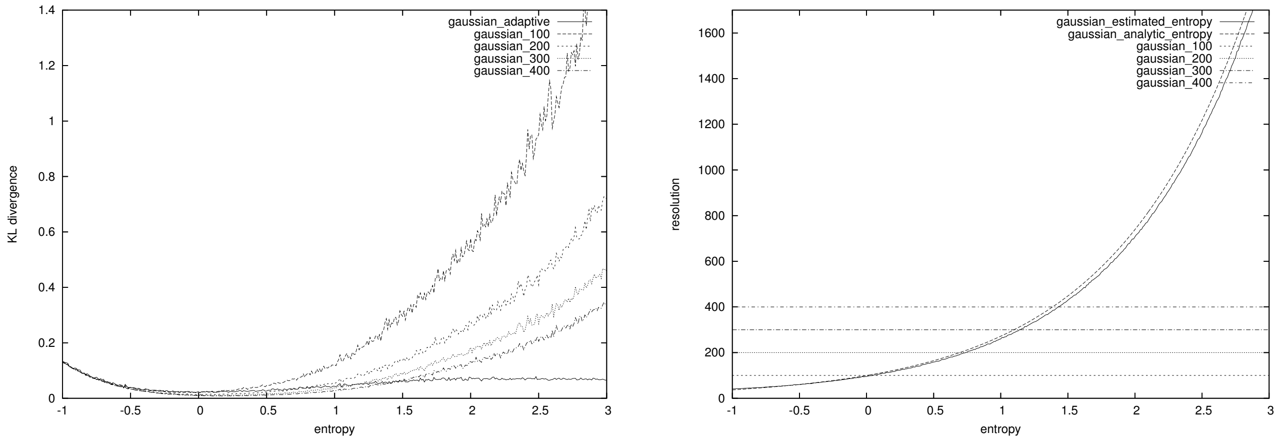

Figure 2. Adaptive sampling on Gaussian target distributions: (a) KL divergence between reconstructed and analytic density as a function of target entropy using constant (n = 100, 200, 300, 400) and adapted resolution; (b) number of samples drawn when using exact and estimated entropy.

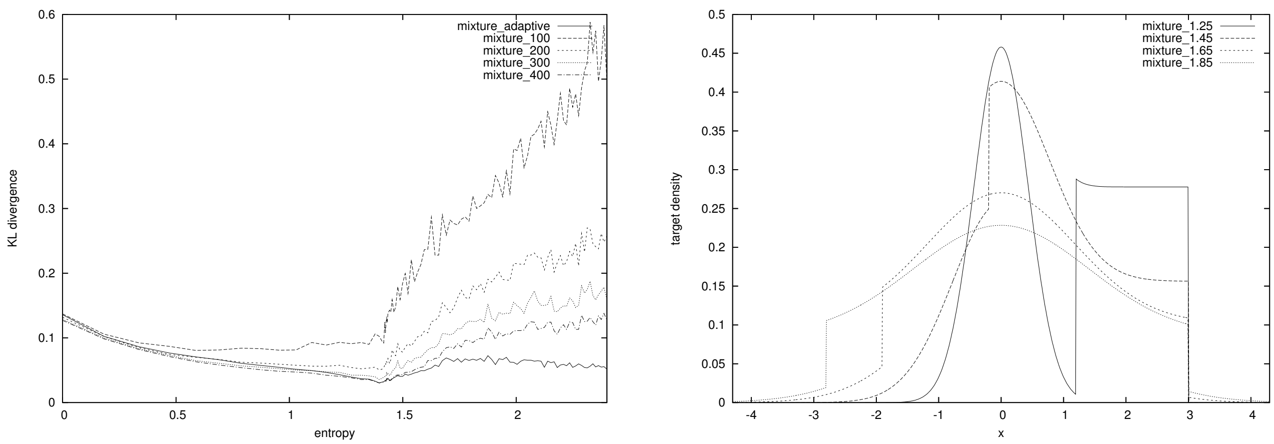

Figure 3. The left plot reports KL divergence for a mixture of Gaussian and uniform distribution. To the right the target distributions for some entropy values are plotted. See text for discussion.

explained by considering the mitigated discontinuities in the target distribution near that value, which are then slightly better accommodated by the smooth Gaussian kernels.

Fig. 4 reports the behavior when approximating a Gaussian q while sampling from a uniform prior p (window width is 10), using four different values for sampling parameter ρ. Given uniformity, it is reasonable to compute the inefficiency factor by $\mathcal{H}(p)/\mathcal{H}(q)$ ([13] proposes an alternative rule). As expected, a denser sampling of the typical set provides higher accuracy. The method draws a quite stable number of particles (namely 980, 730, 480, 220), a feature that complies with the inefficiency of sampling from an uninformative prior.

4.2. Tracking

To illustrate the advantages in a tracking context on a qualitative basis [1] we simulate variable frame rate by dropping a random number (drawn from a Gaussian with mean 2.5 and variance 1.0) of frames after each filter iteration. We use *SmarTrack* [15] to track three people moving in a room, and report statistics about the number of particles used to represent propagated distributions before likelihood update. Fig 5 shows that, when the frame rate is low, more particles are generated since longer prediction times increase system uncertainty. In order to robustly track this sequence using standard sampling, a constant, high number of parti-

[1] Future work will address quantitative assessment, which could be provided by reporting track longevity on a large, annotated database.

prediction	avg. res.	avg. ent.	avg. vol.
67 ms	35	-1.96	0.14
133 ms	58	-1.46	0.23
200 ms	91	-1.01	0.36
267 ms	108	-0.84	0.43
333 ms	133	-0.63	0.53
500 ms	150	-0.51	0.59
567 ms	159	-0.45	0.64

Figure 5. Tracking with variable frame rate (sequence of 3min duration): (a) table showing average belief resolution, entropy and typical set volume for different prediction times, (b) corresponding distribution, and (c) a processed frame with propagated particles and tracker output.

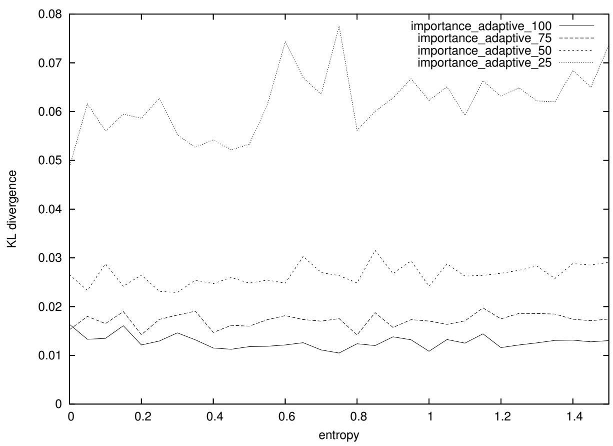

Figure 4. Importance sampling with four different parameter values (ρ = 100, 75, 50, 25).

cles must be generated at each iteration. This unnecessarily slows down the tracker also when the available frame rate is high, possibly introducing additional uncertainty due to the dropping of frames that the tracker in this case is not able to process in real time. This is a critical issue in complex scenes involving multiple occlusions. Sampling a compact while consistent representation is therefore essential.

5. Conclusions

A method for adapting the resolution in particle filters has been proposed. The method builds upon entropy, which directly links to uncertainty, and the AEP theorem to decide on-the-fly the number of particles needed to maintain uncertainty and ambiguity present in sensor data. This way, the trade-off between robustness and computational efficiency is self managed by the filter in a consistent manner, which proves to be an essential capability for the successful real time tracking of complex dynamics and scenes.

References

[1] S. Arulampalam, A. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on–line non–linear/non–gaussian bayesian tracking. *IEEE Trans. Signal Processing*, 50(2):174–188, 2002.

[2] M. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 1991.

[3] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[4] D. Fox. Adapting the sample size in particle filters through KLD-sampling. *Int.l Journal of Robotics Research*, 22(12):985–1004, 2003.

[5] D. Fox, W. Burgard, and S. Thrun. Active Markov localization for mobile robots. *Robotics and Autonomous Systems*, 25(4):195–207, 1998.

[6] M. Isard and J. MacCormick. BraMBLe: A Bayesian multiple-blob tracker. In *IEEE Conf. Computer Vision*, 2001.

[7] P. Jensfelt and S. Kristensen. Active global localisation for a mobile robot using multiple hypothesis tracking. *IEEE Trans. Robotics and Automation*, 17(5):748–760, 2001.

[8] Z. Khan, T. Balch, and F. Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(11):1805–1819, 2005.

[9] D. Koller and R. Fratkina. Using learning for approximation in stochastic processes. In *Conf. Machine Learning*, 1998.

[10] O. Lanz. Approximate Bayesian multibody tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(9):1436–1449, 2006.

[11] C. Olsen. Probabilistic self-localization for mobile robots. *IEEE Trans. Robotics and Automation*, 16(1):55–66, 2000.

[12] J. Porta, B. Terwijn, and B. Krose. Efficient entropy-based action selection for appearance-based robot localization. In *IEEE Conference on Robotics and Automation, ICRA*, 2003.

[13] A. Soto. Self adaptive particle filter. In *Int.l Joint Conf. on Artificial Intelligence*, 2005.

[14] B. Vo, S. Singh, and A. Doucet. Sequential Monte Carlo methods for Bayesian multi-target filtering with Random Finite Sets. *IEEE Trans. Aerospace and Electronic Systems*, 41(4), 2005.

[15] *SmarTrack* - a *SmarT* people *Track*er. Patent pending. Online: http://tev.itc.it/TeV/Technologies/SmarTrack.html.