Probabilistic object tracking for global optimization

M. Nieto and O. Otaegui

VICOMtech, Mikeletegi Pasealekua, 57, Parque Tecnológico, 20009, Donostia-San Sebastián, Spain

ABSTRACT

Object tracking in video-surveillance applications is a major topic in the signal processing and computer vision community. Its aim is to estimate properties of the imaged objects within a scene, such as their dimensions, motion, visual features, etc. Typically, object tracking is treated as an inference problem through probabilistic methods (Bayesian filtering), where the images act as noisy observations of the reality. The nature of inference methods allows abstracting from the specific problem or scenario and modeling very different situations under the same framework. In this field, particle filters (or non-linear, non-Gaussian filters) have emerged as a popular tool to make estimates of posterior distributions within the Bayesian frame. In this work, we review the use of particle filters and exemplify its performance with two typical vision problems: human tracking, and vehicle tracking.

1. INTRODUCTION

Tracking objects may be understood as the process of estimating the state of a dynamic system by analyzing the available observations of the state over time [1]. Probabilistic approaches have arisen as the most powerful methodology to accomplish the detection and tracking of objects in computer vision applications. The reason is that these approaches tackle, in a single and elegant procedure, the noise of the observations, their potential distortions, occlusion between objects, as well as provide the required spatial and temporal coherence to the estimations.

One of the most robust tools designed for object tracking comes from Monte Carlo methods and their sequential version, usually called *particle filters* [1][2][3].

This paper reviews the formulation of the object tracking problem using the Bayesian approach. Two practical examples introduce the use of different alternatives to accomplish object tracking using particle filters. On the one hand, multiple object tracking using importance sampling particle filters [4], and, on the other hand, vehicle tracking using MCMC-based (Markov Chain Monte Carlo) MAP (Maximum A Posteriori) solutions.

1.1. BAYESIAN FILTERING

The target of any tracking system is the estimation of the properties of objects through time. These properties can be related to their magnitude (length, width, height), kinetics (speed, acceleration), or any other characteristic features. As a form of generalization, any object can be defined as a set of random variables at time t, or state-vector, $\mathbf{x}_t$. Bayesian filtering aims to define all the information about this unobservable state provided the set of noisy observations up to time *t*, Z^t, by means of retrieving the posterior density function $p(\mathbf{x}_t|Z^t)$.

The analytical expression of the posterior density can be decomposed using the Bayes' rule as:

$$p(\mathbf{x}_t|Z^t) = kp(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|Z^{t-1})$$

where $p(\mathbf{z}_t|\mathbf{x}_t)$ is the likelihood function that models how likely the measurement $\mathbf{z}_t$ would be observed given the system state vector $\mathbf{x}_t$, and $p(\mathbf{x}_t|Z^{t-1})$ is the prediction information, since it provides all the information we know about the current state before the new observation is available. The constant *k* is a scale factor that ensures that the density integrates to one.

The prediction distribution is given by the Kolmogorov-Chapman equation [1]:

$$p(\mathbf{x}_t|Z^{t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|Z^{t-1})d\mathbf{x}_{t-1}$$

In general, this expression is analytically intractable, and thus, the estimation of the posterior density must be done using numerical approximations through optimization procedures. A noteworthy exception is the existence of an exact analytical solution when both the likelihood function and the dynamic function are linear, corrupted with Gaussian noise. In this specific case, the well known Kalman filter provides the solution to the estimation of the posterior density [5].

For more general problems, where the dynamics of the system are nonlinear, or the noise corrupting the dynamics of the likelihood is not Gaussian, particle filters can be applied as a suboptimal solution to the problem [1].

In general, particle filters work making estimations of the posterior density function $p(\mathbf{x}_t|Z^t)$ as a set of samples, which are propagated through time applying the expected dynamics and corrected given the instantaneous observations. Although there are many existing approaches to perform this approximation, two main groups of techniques are exemplified in the next two sections.

1.2. IMPORTANCE SAMPLING

The first group of particle filters uses importance sampling strategies to propagate samples of the posterior distribution of the target state vector. The key idea is that the posterior density is approximated with a set of random samples (particles) with an associated weight.

Let us hypothesize that the posterior can be expressed as a set of weighted samples

$$p(\mathbf{x}_t|Z^t) \approx \sum_{i=1}^{N_s} \omega_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i)$$

where the weights are chosen according to the principle of importance sampling [1]. In general, this principle states that if there is a probability density from which it is difficult to draw samples, we can instead draw samples from a simpler one, typically called importance density. In order to obtain the desired approximation, the weights must satisfy:

$$\omega_t^i = \frac{p(\mathbf{x}_t^i|Z^t)}{q(\mathbf{x}_t^i|Z^t)}$$

In [1] it is demonstrated, after some operations that this leads to the following recursive expression of the weights:

$$\omega_t^i \propto \omega_{t-1}^i \frac{p\left(\mathbf{z}_t|\mathbf{x}_t^i\right)p(\mathbf{x}_t|\mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, Z^t)}$$

Typically, and for the sake of simplicity, the proposal density is chosen to be the same as the dynamics of the sate vector, $q\left(\mathbf{x}_t^i\middle|\mathbf{x}_{t-1}^i, Z^t\right) = p(\mathbf{x}_t|\mathbf{x}_{t-1}^i)$, such that the expression of the weights is limited to consider the likelihood function of the particles:

$$\omega_t^i \propto \omega_{t-1}^i p\left(\mathbf{z}_t|\mathbf{x}_t^i\right)$$

Besides, when using resampling strategies, the weights are normalized such that $\sum_{i=1}^{N_s} \omega_t^i = 1$, and thus

$$\omega_t^i \propto p\left(\mathbf{z}_t|\mathbf{x}_t^i\right)$$

This filter is typically known as sequential importance re-sampling (SIR), and is the basis for numerous works that track objects [3][4].

1.3. MCMC

Let us now hypothesize that the posterior can be expressed as a set of unweighted samples

$$p(\mathbf{x}_{t-1}|Z^{t-1}) = \frac{1}{N_s}\sum_{i=1}^{N_s} \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^{\mathrm{i}})$$

then

$$p(\mathbf{x}_t|Z^{t-1}) \approx \frac{1}{N_s}\sum_{i=1}^{N_s} p(\mathbf{x}_t|\mathbf{x}_{t-1}^i)$$

Therefore we can directly sample from the posterior distribution since we have its approximate analytic expression:

$$p(\mathbf{x}_t|Z^t) \propto p(\mathbf{z}_t|\mathbf{x}_t) \sum_{i=1}^{N_s} p(\mathbf{x}_t|\mathbf{x}_{t-1}^i)$$

For this purpose we need a sampling strategy, like the Metropolis-Hastings (MH) algorithm, which dramatically improves the performance of traditional particle filters based on importance sampling. As a summary, the MH generates a new sample according to an acceptance ratio that can be written as:

$$\alpha = \frac{p(\mathbf{x}_t^j|Z^t)}{p(\mathbf{x}_t^{j-1}|Z^t)}\frac{q(\mathbf{x}_t^{j-1}|\mathbf{x}_t^j)}{q(\mathbf{x}_t^j|\mathbf{x}_t^{j-1})}$$

where j is the index of the samples of the current chain. The proposed sample $\mathbf{x}_t^j$ is accepted with probability $min(\alpha, 1)$. If the sample is rejected, the current state is kept, i.e. $\mathbf{x}_t^j = \mathbf{x}_t^{j-1}$. The proposal density $q()$ might be any function from which it is easy to draw samples. Typically it is chosen as a normal distribution, which is symmetric, i.e. $q(\mathbf{x}_t^j|\mathbf{x}_t^{j-1}) = q(\mathbf{x}_t^{j-1}|\mathbf{x}_t^j)$ such that its corresponding term can be removed from the acceptance ratio expression.

Besides, it is a common practice to select a subset of samples from the chain to reduce their correlation and to discard a number of initial samples to reduce the influence of initialization. Therefore, to obtain N_s effective samples of the chain it is required to generate a total number of samples $N = B + cN_s$, where B is the number of initial samples, and c is the number of samples discarded per valid sample.

2. MULTIPLE OBJECT TRACKING – HUMAN TRACKING

In this section we describe a variant of the SIR filter that can handle multiple objects in a single filter. The basic difference of the proposed approach with respect to traditional particle filter methods [3] is that it includes a measurement-guided clustering procedure that divides the set of particles into M subsets, each one representing one object of the scene. The set of particles is extender by adding an index for each particle, c_t^i, where $c_t^i \in \{1,2, \ldots, M\}$.

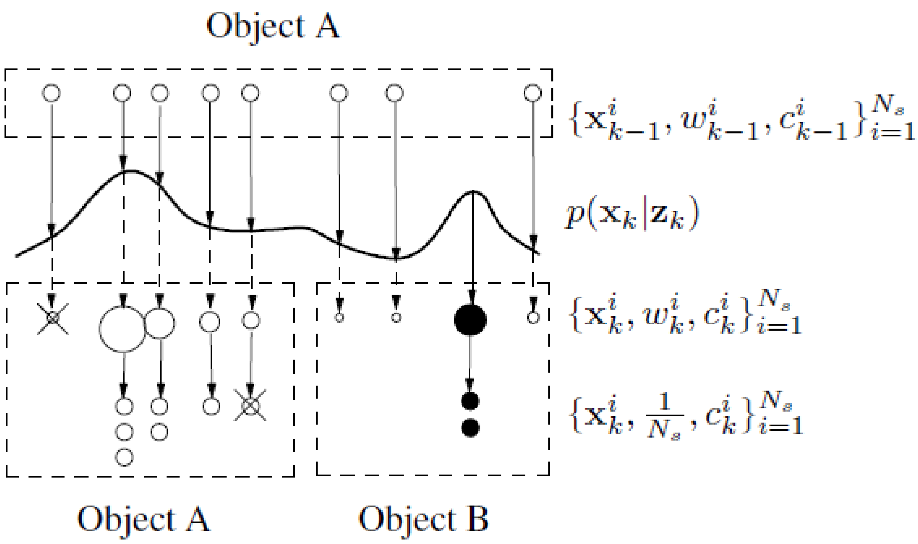

Figure 1: SIR filters tend to concentrate particles around main modes of the distribution. Appropriately injecting new particles (painted in black) resamples the space and allows multiple modes to be tracked. Multiple objects are detected by the consequent spatial reclustering of particles.

Figure 1depicts the main idea of the problem: typically, after an arbitrary number of time instants, most particles may have been grouped around the regions with higher density values. The number of particles in other regions of the considered state-space may have been correspondingly reduced to the point of missing other objects if their likelihood is excessively narrow (as in the right mode shown in the figure), or insufficiently relevant. The resampling step, though reducing the degeneracy problem, usually worsens this missing-object situation.

The proposed algorithm works injecting a new particle (the black one in the figure) near this missing mode substituting one of the low-weighted particles. The result is an improved set of particles that better represents the multimodal posterior density, and thus allows appropriate spatial reclustering that deytermines the values of c_t^i. This is handled through the definition of the novelty distribution, $n(\mathbf{z}_t)$, which handles the reclustering of particles into sub-sets.

2.1. Likelihood model for Human Tracking

This section considers the problem of tracking multiple objects, namely humans, in typical surveillance indoor scenarios, where a previous segmentation of the images is available [4]. To show the performance of the proposed particle filter, we are using the segmentation described in [7].

The objects are represented by ellipses, as shown in , although any other model might be used. The state-vector is then defined as $\mathbf{x}_t = (x, y, \dot{x}, \dot{y}, a, b)^T$, where the first two items are the position of the center of the ellipse, in pixel units, denoted as $\mathbf{p}_t = (x, y)^T$, the third and fourth are its velocity, and the two last ones are the axes of the ellipse. For the sake of simplicity we will not consider rotation on the model, introducing no lack of accuracy in the considered scenarios. The measurement vector is composed by J_t pixels that have been classified as belonging to moving objects by the segmentation process: $\mathbf{z} = \{x_j, y_j\}_{j=1}^{\mathrm{J}_t}$.

The function that evaluates the novelty of the measurements is defined as

$$\mathrm{h}(\mathbf{z}_\mathrm{t}^\mathrm{j}|\mathbf{x}_\mathrm{t}) \propto \exp\left(-\frac{1}{2}(\mathbf{z}_\mathrm{t}^\mathrm{j} - \mathbf{p}_\mathrm{t})^\mathrm{T} \mathrm{C}^{-1} (\mathbf{z}_\mathrm{t}^\mathrm{j} - \mathbf{p}_\mathrm{t})\right)$$

which is the normal distribution of the Mahalanobis distance between the measurement and the position of the particle. *C* is the covariance matrix that normalizes dimensions.

Let $Z_t^m = \{\mathbf{z}_t^j\}_{j \in M}$ be the set of measurements associated to the object m. Then, the likelihood distribution for this object can be expressed as:

$$\mathrm{L}(Z_t^m|\mathbf{x}_\mathrm{t}) \propto \frac{\mathrm{N_{in}}}{\pi \mathrm{ab}} \frac{\mathrm{N_{in}}}{\mathrm{N}} \prod_{j \in M} \left(\rho + \mathrm{h}(\mathbf{z}_\mathrm{t}^\mathrm{j}|\mathbf{x}_\mathrm{t})\right)$$

where $\mathrm{N_{in}}$is the number of measurements associated to the object m that fall inside the ellipse $\mathbf{x}_\mathrm{t}$ under analysi, and N is the total number of measurements for the object *m*. The two external factors of the product represent, respectively, the relation between the measurements inside the ellipse with respect to the area of the ellipse, and with respect to the total number of measurements. These global factors tend to give high likelihood values to those ellipses best fitting the set of measurements. The product just combines the contributions of independent measurements, using a regularization factor, ρ, that works as a base level that avoids outlier measurements to affect the overall likelihood value of the particle. An example of the behavior of this function is shown in Figure 3.

Figure 2 : Visual tracking: (a) original images; (b) segmentation; and (c) set of particles (blue ellipses), and the best estimation, in magenta.

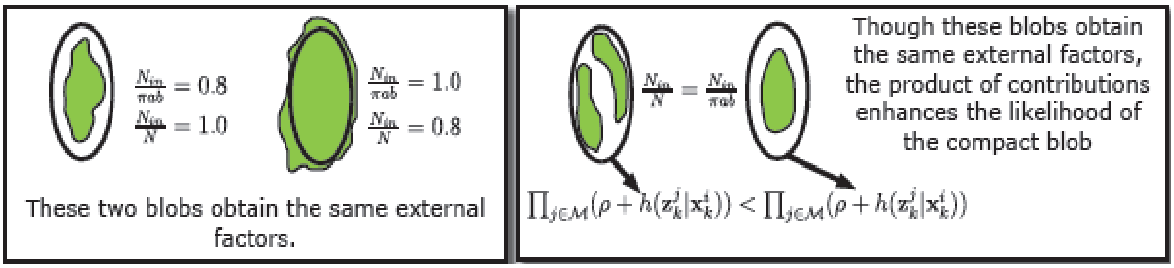

Figure 3: Likelihood function for ellipses provided the pixels classified as foreground (in green).

2.2. RESULTS

The proposed particle filter was tested on different surveillance-like example video sequences where images size was 360x288 pixels. The initialization of the system is done automatically (at the beginning of the sequence and when there are no measurements in the image): the particles are initialized such that their centers follow a uniform distribution covering the entire image, setting their velocities as a normal distribution around zero. The axes of the ellipse are set randomly according to the expected range of sizes of the objects (remark that the perspective effect makes the objects to appear with different sizes). The experiments have shown that using this segmentation and the proposed likelihood models, a reduced set of particles is enough to quite accurately detect and track multiple objects. Good results have been obtained using 100 particles for tracking between1 and 4 objects (examples are shown in Figure 4). One key issue of the proposed strategy is that automatically adapts the number of particles to be used for each object, according to the measurements: a new object receives a single particle the first time it appears, and subsequentely, the reclustering step assigns more particles until the measurements associated to the object get low novelty values.

Figure 4: Examples of the application of the particle filter for multiple object tracking.

3. VEHICLE TRACKING

The target of this second version of the particle filter is the estimation of the dimensions of a detected and tracked vehicle, which is modeled as a rectangular cuboid, defined by its width, height and length, $\mathbf{x}_t = (w_t, h_t, l_t)^T$, in order to classify it as one of a set of predefined classes. The estimation is done for each time instant, *t*, based on the previous estimations and the new incoming image observations.

3.1. LIKELIHOOD MODEL

For each new image, the observations are the bounding boxes of the 2D silhouettes of the detected vehicles projected into the rectified view of the road plane. Considering the cuboid model of the vehicle, and that the yaw angle can be considered approximately zero, we can reproject a 3D ray from the far-most corner of the 2D bounding box and the optical center.

There are infinite points on this ray that are projected in the same image point and therefore correspond to a putative solution to the parameters of the cuboid. Figure 5 illustrates this projective ambiguity. Nevertheless, there are a number of constraints that bound the solution to a segment of the ray: positive and minimum height, width and length.

Therefore, the likelihood function may be any function that fosters volume hypotheses near the reprojection ray. For the sake of simplicity, we choose a normal distribution on the point-line distance. The covariance of the distribution expresses our confidence about the measurement of the 2D silhouette and the calibration information. The likelihood function can be written as

$$\mathrm{p}(\mathbf{z}_\mathrm{t}|\mathbf{x}_\mathrm{t}) \propto \exp((\mathbf{y}_\mathrm{t} - \mathbf{x}_\mathrm{t})^\mathrm{T} \mathrm{S}^{-1} (\mathbf{y}_\mathrm{t} - \mathbf{x}_\mathrm{t}))$$

where $\mathbf{x}_\mathrm{t}$ is a volume hypothesis, and $\mathbf{y}_\mathrm{t}$ is its projection onto the reprojection ray. The position of $\mathbf{y}_\mathrm{t}$ can be computed from $\mathbf{x}_\mathrm{t}$ as the intersection of the ray and a plane passing through $\mathbf{x}_\mathrm{t}$ and whose normal vector is parallel to the ray. For that reason it is convenient to represent the ray as a Plücker matrix $\mathrm{L}_\mathrm{t} = \mathbf{a}\mathbf{b}^\mathrm{T} - \mathbf{b}\mathbf{a}^\mathrm{T}$, where **a** and **b** are two points of the line, e.g. the far-most point of the 2D silhouette, and the optical center, respectively. These two points are expressed in the WHL coordinate system. For this purpose, and provided the calibration of the camera is available, we need a reference point of the 2D silhouette. We have observed that the point with less distortion is typically the closest point of the quadrilateral to the optical center, whose coordinates are $\mathrm{X}_{\mathrm{t},0} = (\mathrm{x}_{\mathrm{t},0}, 0, \mathrm{z}_{\mathrm{t},0})^\mathrm{T}$ in the XYZ world coordinate system. This way, any XYZ point can be transformed into a WHL as $\mathbf{x}_t = R_0 X_t - X_{t,0}$. Nevertheless, the relative rotation between these systems can be approximated to the identity, since the vehicles typically drive parallel to the defined OZ axis.

The plane is defined as $\pi_t = (\mathbf{n}_t^T, D_t)^T$, where $\mathbf{n}_t = (n_x, n_y, n_z)^T$ is the normal to the ray L_t, and $D_t = -\mathbf{n}_t^T \mathbf{x}_t$. Therefore, the projection of the point on the ray can be computed as $\mathbf{y}_t = L_t \pi_t$.

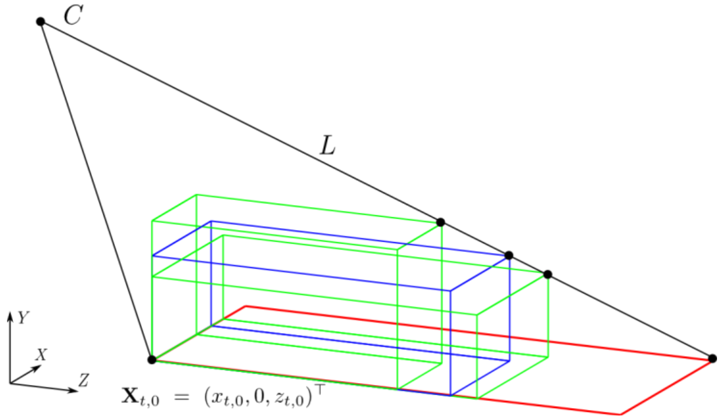

Figure 5: Projective ambiguity: a given 2D observation in the OXZ plane (in red) of a true 3D cuboid (blue) may also be the result of the projection of a family of cuboids (in green) with respect to camera *C*.

3.2. PRIOR MODELS

The information about the volume of the vehicle can be encoded as the product of two functions, each one modeling two independent sources of information:

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, M) = p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_t | M)$$

where $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ represents the dynamic model of the system. In our case, we will assume that a vehicle is a non-deformable rigid object, such that it does not vary its dimensions through time, and thus

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) \propto \exp((\mathbf{x}_t - \mathbf{x}_{t-1})^T (\mathbf{x}_t - \mathbf{x}_{t-1}))$$

The second term of the prior model, $p(\mathbf{x}_t | M)$, contains the information that we have about typical configurations of vehicle dimensions, i.e. typical proportions of vehicles according to a number of models, such as truck, motorcycle, car, etc. Let us represent this information as a set of clusters that can be parameterized as a mixture of normal distributions in the WHL space: $M = \{\mathbf{x}_m\}_{m=1}^{M}$. Therefore,

$$p(\mathbf{x}_t | M) = \sum_{m=1}^{M} p(\mathbf{x}_t, \mathbf{x}_m)$$

where $\mathbf{x}_m = (W_m, H_m, L_m)^T$ and

$$p(\mathbf{x}_t, \mathbf{x}_m) \propto \exp((\mathbf{x}_t - \mathbf{x}_m)^T S_m^{-1} (\mathbf{x}_t - \mathbf{x}_m))$$

Table 1: Example configuration of vehicle models

Vehicle type	W_m	H_m	L_m	σ_w	σ_h	σ_l
Car	1.6	1.5	4	0.1	0.1	0.2
Motorbike	1.6	1.5	2	0.1	0.1	0.2
Truck	2.0	2.5	7	0.2	0.3	1.0
Trailer	1.6	1.5	7	0.1	0.1	2.0
Bus	2.0	2.5	10	0.2	0.3	1.0

Table 1 exemplifies a set of vehicle models. The Gaussian model ensures that the vehicle models are not rigid nor fixed, in contrast with typical wireframe models, and thus enhances the flexibility of prior information. For instance, trucks can be modeled as a 3D Gaussian centered at (2.0, 2.5, 7) with high variance values, since trucks may vary significantly in length or height.

3.3. ALGORITHM COMPLEXITY REDUCTION

Once we have defined the prior and observation models, the complete expression of the MH acceptance ratio is given by:

$$\alpha = \frac{\mathrm{p}(\mathbf{z}_\mathrm{t}|\mathbf{x}_t^j)}{p(\mathbf{z}_\mathrm{t}|\mathbf{x}_t^{j-1})} \frac{\sum_{i=1}^{N_s} p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^i)}{\sum_{i=1}^{N_s} p(\mathbf{x}_t^{j-1}|\mathbf{x}_{t-1}^i)} \frac{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^j, \mathbf{x}_\mathrm{m})}{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^{j-1}, \mathbf{x}_\mathrm{m})}$$

By drawing N_s effective samples with the MH algorithm we have the target approximation of the posterior distribution. Hence, we can compute point-estimates of the state vector $\mathbf{x}_t$ and thus estimate the volume of the 3D cuboid at each time instant. For instance we can use the sample mean as the simplest statistic, which is valid enough since the posterior distribution can be assumed to be unimodal without loss of generalization.

Nevertheless, the generation of the Markov chain implies a significant amount of computations, since the computational complexity is $O(NN_s)$. The reason is that for each proposed sample $\mathbf{x}_t^j$, the complete set of previous samples $\{\mathbf{x}_t^j\}_{i=1}^{N_s}$ has to be evaluated to compute the acceptance ratio.

To reduce to linear time operation, i.e. $O(N)$, we can instead select a single previous sample, $\mathbf{x}_{t-1}^*$, from the set. Khan et al. [8] propose to select a random sample from the set, although we have observed much better performance selecting the most likely sample of the set, or the point-estimate of the previous time instant. The acceptance ratio expression is then simplified to:

$$\alpha = \frac{\mathrm{p}(\mathbf{z}_\mathrm{t}|\mathbf{x}_t^j)}{p(\mathbf{z}_\mathrm{t}|\mathbf{x}_t^{j-1})} \frac{p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^*)}{p(\mathbf{x}_t^{j-1}|\mathbf{x}_{t-1}^*)} \frac{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^j, \mathbf{x}_\mathrm{m})}{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^{j-1}, \mathbf{x}_\mathrm{m})}$$

Regarding the specific nature of our problem, and additional great reduction of the complexity of the sampling step can be achieved if we force the samples to belong to the ray defined by the likelihood model. This is equivalent to reduce the problem to a one-dimensional search on the ray. On the one hand, the proposal density can be now defined as a one-dimensional normal distribution that draws samples on the ray, as well as the dynamic model. Therefore, the samples are now drawn based on the simplified expression of the acceptance ratio:

$$\alpha = \frac{p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^*)}{p(\mathbf{x}_t^{j-1}|\mathbf{x}_{t-1}^*)} \frac{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^j, \mathbf{x}_\mathrm{m})}{\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}(\mathbf{x}_t^{j-1}, \mathbf{x}_\mathrm{m})}$$

subject to $\mathbf{x}_t^j \in L_t$.

The implementation of the algorithm can be as well simplified if the state space is reduce to a discrete number of states, namely $\{\mathbf{y}_m\}_{m=1}^M$, i.e. the projections of the vehicle models on the observed ray. Under this assumption, the algorithm computes the posterior probability of each $\mathbf{y}_{t,m} = L_t\pi$ as proportional to $p(\mathbf{y}_{t,m}|\mathbf{y}_{t-1}^*)\mathrm{p}(\mathbf{y}_{t,m}, \mathbf{x}_\mathrm{m})$, and determines the point-estimate of $p(\mathbf{x}_t|Z^t)$ as the most likely projection $\mathbf{y}_{t,m}$.

Figure 6: Example results of 3D vehicle modeling, including different size vehicles and type of perspective.

3.4. RESULTS

The proposed system overcomes the problems of 2D strategies that aim to measure the dimensions of the vehicles for classification purposes in perspective images. In order to evaluate the performance of the vehicle classification, we have tested the proposed solution for a set of videos of different roads and perspectives, with an aggregate duration of more than 5 hours. The total number of detected vehicles in the video sequence is 2551/2585 (98,7%). The target application required the classification of vehicles into two broad categories: light and heavy. Considering the detected vehicles, the system correctly classified 2214/2248 light vehicles, and 337/337 heavy vehicles according to their volume. Some example images of the renderization of the estimated 3D model are shown in Figure 6. As shown, in most situations, the cuboid fits approximately the volume occupied by the vehicles (with some inaccuracy due to insufficient perspective distortion or excessively long vehicles), and thus allow to classify vehicles in the required categories.

4. CONCLUSIONS

This paper introduces concepts related to object tracking using Bayesian inference methods. Particle filters are analyzed as their importance is continuously increasing within the scientific community to solve complex dynamic problems. In particular, human and vehicle tracking is presented as example applications for visual tracking. On the one hand, human tracking is solved using a variant of the well-known SIR algorithm, whereas vehicle tracking is faced from the sequential sampling perspective of the Monte Carlo methods. In both cases, the results are excellent both in detection rates and fitting accuracy.

REFERENCES

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp: "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, 2002.

[2] A. Doucet, S. Godsill, and C. Andrieu: "On sequential Monte Carlo sampling methods for Bayesian filtering", Statistics and Computing, vol. 10, no. 3, pp. 197-208, 2000.

[3] M. Isard and A. Blake: "CONDENSATION – conditional density propagation for visual tracking", International Journal on Computer Vision, vol. 29, no. 1, pp. 5-28, 1998.

[4] M. Nieto, C. Cuevas, L. Salgado, "Measurement-based reclustering for multiple object tracking with particle filters", IEEE Int. Conf. on Image Processing, ICIP 2009, Cairo, Egypt, pp. 4097-4100, 7-10 Nov. 2009. (DOI 10.1109/ICIP.2009.5413709).

[5] P. S. Maybeck, "Stochastic models, estimation and control," Academic Press, 1979.

[6] D. Serby, E.-K. Meier, and L. Van Gool, "Probabilistic object tracking using multiple features," in IEEE Proc. International Conference on Pattern Recognition, vol. 2, pp. 184-187, 2004.

[7] C. Cuevas, N. García, and L. Salgado, "A new strategy based on adaptive mixture of gaussians for real-time moving objects segmentation," in SPIE Real-Time Image Processing, vol. 6811, pp. 681111-1-12, 2008.

[8] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 11, pp. 1805-1819, 2005.