A Tutorial for Rao-Blackwell Dimension Reduction in Visual Tracking Applications

Xinyu Xu and Baoxin Li
Department of Computer Science and Engineering
Arizona State University

In many inference problems based on a multi-dimensional state space model, the computation cost may explode and the estimation accuracy may deteriorate rapidly with the increase of the dimensionality of the state space, rendering many conventional techniques ineffective in face of a high dimensional state space model. For example, particle-filter-based methods become quite inefficient when being applied to a high dimensional state space since a prohibitively large number of samples may be required to approximate the underlying density functions with desired accuracy. An effective solution to solving this problem is to use the Rao-Blackwell theorem in reducing the dimension of the state vector that needs to be estimated. This short tutorial illustrates the key idea of such an approach by using visual tracking in surveillance as an example.

The key idea of Rao-Blackwell dimension reduction is to make use of the "structural" information inherent in the problem itself to analytically infer part of state parameters conditional upon other state components which will be estimated by the Sequential Monte Carlo approaches. This directly leads to increased estimation accuracy. For the visual tracking problem, let us denote the state to be estimated as X_t and observation as Z_t with subscript t the time index. If we can partition the original state-space X_t into two parts R_t (*root variables*), and L_t (*leaf variables*), such that the distribution of leaf variables can be computed analytically given the root variables (this is equal to say the distribution $p(L_{1:t}|R_{1:t},Z_{1:t})$ can be estimated analytically) , therefore an approximation of $p(R_{1:t}|Z_{1:t})$ using Monte Carlo methods yields straightforwardly an approximation of joint posterior $p(R_{1:t},L_{1:t}|Z_{1:t})$. The justification for this decomposition follows from the factorization of the probability:

$$p(R_{1:t},L_{1:t}|Z_{1:t})=p(L_{1:t}|R_{1:t},Z_{1:t})p(R_{1:t}|Z_{1:t}) \tag{1}$$

The author in [1] pointed out that $p(L_{1:t}|R_{1:t},Z_{1:t})$ can be efficiently updated using Kalman filter when the initial uncertainty for leaves is Gaussian, and the CPDs (Conditional Probability Distributions) of the observation model and system dynamics for leaves are linear functions of the leaf states. In our work, we proposed Rao-Blackwellised particle filter for tracking, with Kalman filter being applied to estimate the leaf states (linear dynamics) and particle filter to estimate the root states (nonlinear parts).

One advantage of the Rao-Blackwell theorem is that it can greatly reduce the variance of the state estimates. Generally, suppose we have an estimator $\xi(R, L)$ depending upon two variables, its variance satisfies [2]:

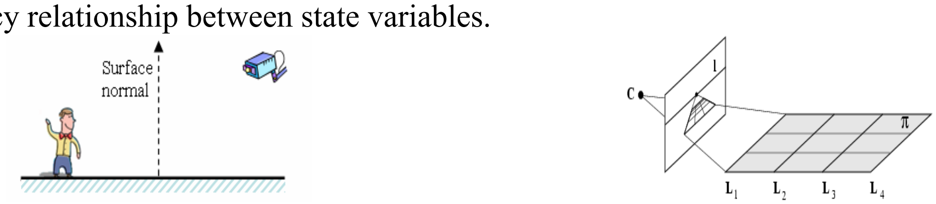
$$Var(\zeta(R,L)) = Var[E\{\zeta(R,L) \mid R\}] + E[Var\{\zeta(R,L) \mid R\} \tag{2}$$

So the variance of estimator $\zeta^{'}=E(\zeta(R,L)|R)$ improves upon the crude estimator $\zeta(R,L)$ by a non-negative term $E[Var\{\zeta(R,L)|R\}]$ [2]. The formal justification can be found in [3]. One can interpret the Rao-Blackwell Theorem by saying that the estimator obtained by the calculation of conditional expectation is superior to the original one.

Two key problems arise when applying RBPF: first is how to identify the conditional relationship between various state variables such that the state variables can be meaningfully partitioned into two (or more) groups; the second is what kind of analytical filter should be used to efficiently estimate the leaf variables conditional on the root variables. In one of our work, we are able to utilize the constraints imposed by typical camera configuration to partition the original state space into two sub-spaces, and Rao-Blackwellised particle filter is then proposed. The major idea is shown below.

Partition the state space

In typical surveillance applications, for most of the time, the tracked objects are constrained to move on a dominant plane (e.g. the ground), and the camera is usually higher than the tracked object. Fig. 1 illustrates such a scene-camera configuration, where (b) is a geometric representation of (a). In Fig. 1(b), suppose a person is moving on the ground plane π, and the ground is projected onto the image plane by camera C, l is the vanishing line for the ground plane. Any scene point projecting onto the vanishing line l is at the same distance from plane π as the camera center [4]. If a scene point is farther from π than the camera is, then its image lies 'above' the vanishing line; and 'below' if it is closer to the ground than the camera [4]. So if the to-be-tracked object is not higher than the camera to the ground, the image of the object will always lie 'below' the vanishing line, and when it moves towards the camera, the size (or scale) of the object on the image will get bigger as the *y* coordinate on image plane gets bigger, and vice versa. From Fig. 1(b), we may see the dependence of the *scale change* of the object on the *translational motion* (or the *y* location in the image domain). In these situations, the constraints due to the camera-scene configuration can be exploited to deduce the dependency relationship between state variables.



(a) **(b)**

Fig. 1. (a) Illustration of a typical camera configuration in surveillance. (b) The dominant plane is projected onto the image plane. If the person moves for example from l_4 to l_1,

then on the image, we may observe that as the y-coordinate gets larger and larger, the size of the person is becoming bigger and bigger.

Formally, in our work we use an 8-D ellipse model to describe the dynamics of the tracked object, as in [5]:

$$\{x, \tilde{x}, y, \tilde{y}, H_y, \widetilde{H_y}, H_x, \widetilde{H_x}\}$$

With the above idea, the scale change of a moving object is related to its position along the *y-axis* (i.e., the vertical axis in the image domain), as illustrated in Fig. 1(b). This facilitates the partition of the original 8-D state space into two groups: the root variables R containing the motion information (including location and velocity), which will be sampled by a particle filter, and the leaf variables L consisting of the scale parameters (including the rate of scale change), which may be estimated by an exact filter. These two groups are denoted by:

$$R = \{x, \tilde{x}, y, \tilde{y}\} \qquad L = \{H_y, \widetilde{H_y}, H_x, \widetilde{H_x}\}$$

The Rao-Blackwellised Particle Filter

The Rao-Blackwellised Particle Filter is shown in Fig. 3.

Input:: $S_{t-1} = \{< R_{t-1}^i, \mu_{t-1}^i, \sigma_{t-1}^i >| i = 1 \cdots N\}$ and Z_t.

For $i = 1{:}N$ do

1.Propagate samples

a) ***Sample object motion***:

$$R_t^{i-} \sim p(R_t \mid R_{t-1}^i, Z_t) = p(Z_t \mid R_t) p(R_t \mid R_{t-1}^i)$$

b) ***Kalman prediction***:

$$L_t^{i-} \sim p(L_t \mid R_t^i, R_{t-1}^i, L_{t-1}^i, Z_t)$$

2.Evaluate weight

a) ***Compute the color histogram.***

b) ***Compute the gradient.***

c) ***Compute the weight:***

$$w_t^i \propto p(Z_t^{CH} \mid R_t^i, L_t^i) \bullet p(Z_t^G \mid R_t^i, L_t^i)$$

End for loop

3.Select samples.

For $i = 1{:}N$ do

4.***Kalman update:***

End for loop

5.***Compute the mean state***. $E[S_t] = \sum_{i=1}^{N} w_t^i s_t^i$

Figure 3: The Rao-Blackwellised particle filter for tracking.

For more discussion on this topic please refer to [6].

References

[1] K. Murphy and S. Russell, “Rao-Blackwellised particle filtering for dynamic Bayesian networks,” in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. Freitas, and N. Gordon, Eds. New York: Springer-Verlag, 2001, chapter 24.

[2] G. Casella, C. P. Robert, “Rao Blackwellisation of Sampling Schemes,” *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.

[3] N. Mukhopadhyay, *Probability and Statistical Inference*, New York: Marcel Dekker, Inc., pp. 359, 2000.

[4] R. Hartley and A. Zisserman*, Multiple View Geometry in Computer Vision,* second edition, Cambridge University Press, March 2004, pp.220.

[5]K. Nummiaro, E. Koller-Meier, and L. V. Gool, “Object tracking with an adaptive color-based particle filter,” in Proc. Symposium for Pattern Recognition of the DAGM, pp. 591-599, 2003.

[6] Xinyu Xu, Baoxin Li, “Rao-Blackwellised Particle Filter for Visual Tracking with Application in Video Surveillance”, VS-PETS Workshop (joint with ICCV 2005), October 15-16, 2005, Beijing.