ON UNCERTAINTIES, RANDOM FEATURES AND OBJECT TRACKING

Vijay Badrinarayanan, Patrick Perez, Francois Le Clerc and Lionel Oisel*

Thomson Corporate Research, Rennes, France,*IRISA-INRIA Rennes, France

ABSTRACT

Algorithms for probabilistic visual tracking hypothesize a distribution of the target state (location, scale, etc.) at every tracking step with an associated information content or equivalently, an uncertainty. One measure of this uncertainty is the differential entropy. In this paper, we present a unified way to approximate the differential entropy of tracking distributions, which then makes it suitable, among other factors, for a qualitative assessment of both deterministic and sequential Monte Carlo simulation based tracking algorithms. We then illustrate the usefulness of this assessment measure via tracking an object by choosing a set of randomly picked features on it, each individually tracked, removed according to an uncertainty analysis and replaced randomly, without any aid of a feature selection algorithm as in current use.

***Index Terms*—** Visual Tracking, Probabilistic Filtering, Differential Entropy, Randomized features

1. INTRODUCTION

Any visual tracking algorithm, at every tracking step, in one way or another, relies on a search mechanism. The search is either deterministic, centered on the previous tracked state (location, scale, velocity or other parameters that make up the state vector), as for instance, in template based tracking or a random walk conditioned on the previous tracked state, as in sequential Monte Carlo based approaches. At each point in the search space, a quantitative measure of similarity is evaluated, which primarily compares the proximity of the search point to the target in an appearance or descriptive space. This ensemble of search points, together with their associated similarity measure or weight, forms the distribution of the tracked state (the term tracking distribution is also used interchangeably). From this distribution, a statistical inference is drawn about the tracked state.

Each distribution can be associated with quantitative information content. This information is the measure of uncertainty about the tracked state. The behavior of this uncertainty measure, over the series of tracking steps, is indicative of the tracking performance. Intuitively, increasing uncertainties over the tracking steps indicate that it is with low confidence that the tracking algorithm infers the tracked state from the hypothesis.

Several intuitive measures of uncertainty can be drawn from a distribution. The modal point is ubiquitous [1], but nevertheless invariant to the spatial structure (uniformity and spatial extent) of the distribution. The variance or covariance [2], is another measure, frequently employed to quantify the effective spatial extent of the hypothesis, but except in a few cases (the Gaussian distribution is an example), they are not indicative of the *non-Gaussianity* [3] of the distribution. For example, a uniform distribution and a Gaussian distribution can be made to have the same variance (covariance) over the same finite support. The variance measure is unable to capture the form differences in the distributions. But intuitively, the uniform distribution is more uncertain than the Gaussian, as a consequence of its flatness. It is clear then, that the uncertainty measure should be indicative of both the form and spatial extent of the distribution. This leads us to investigate an information theoretic quantity.

The information-theoretic quantity we primarily investigate as an indicator of tracking performance is the *differential entropy* of the tracked state [4] (The state in this paper is taken to be the location of the object). This choice is motivated by two reasons. First, this quantity is a measure of the spatial structure of a probability distribution [3]. Secondly, it is amenable to usage, via tractable approximations, in a probabilistic filtering framework and importantly, it can also quantify information exchange for fusion algorithms.

Before the differential entropy of the tracked state can be measured in either class of trackers, it is important to see how deterministic tracking may be viewed in a probabilistic filtering framework, which practically allows the two classes of trackers to be analyzed uniformly. We dedicate Section 2 of this paper to this question. Building on this, in Section 3, we derive approximations to the differential entropy of the tracked state using the filtering framework. With these approximations, in Section 4 we construct a stand-alone object tracker influenced by its performance analysis. We term this tracker the *randomized feature-tracker*. We provide experimental results of tracking in complex scenes and discuss the differentiating properties of this tracker in Section 5. We draw conclusions in Section 6.

2. DETERMINISTIC TRACKING: A PROBABILISTIC VIEWPOINT

A tracker, in either of the two broad classes, can be modeled as a processing box at every tracking step with data and other priors at the input and a resulting tracking distribution at the output. The tracking distribution in probabilistic tracking is the posterior distribution, while in deterministic tracking, it could, for instance, be a normalized correlation surface arising from template matching or a distribution of feature matched points and match weights in a feature matching scheme. The key idea behind the unification principle is to bring about a way to cast characteristically different tracking distributions as approximations to probabilistic posterior distributions. We discuss this idea below.

The following definitions will be frequently encountered in the rest of this paper.

$x_{0:k}$: The tracked state trajectory up to tracking step k. The state here is the location of the target on the two dimensional image grid.

$z_{1:k}$: The data or image sequence up to tracking step k .

$p\left(\frac{x_{0:k}}{z_{1:k}}\right)$: The posterior distribution of the tracked state given the data up to time k.

In Sequential Monte Carlo (SMC) approach, the posterior distribution is approximated by a set of samples and corresponding weights $\{x^i_{0:k}, w^i_k\}^N_{i=1}$ as follows.

$$p\left(\frac{x_{0:k}}{z_{1:k}}\right) \approx \Sigma^N_{i=1} w^i_k \delta\left(x_{0:k} - x^i_{0:k}\right) \quad (1)$$

Assuming a first order Hidden Markov Model (HMM) structure for the state space model, the principle of *importance sampling* is invoked to derive a recursive form of weight evaluation as given below [5].

$$w_k^i \propto \frac{p\left(\frac{z_k}{x_k^i}\right) p\left(\frac{x_k^i}{x_{k-1}^i}\right)}{q\left(\frac{x_k^i}{x_{k-1}^i, z_{1:k}}\right)} w_{k-1}^i, \quad (2)$$

where, $q\left(\frac{x_k^i}{x_{k-1}^i, z_{1:k}}\right)$ is the *importance sampling density*.
This weight recursion in eqn.(2) is the starting point to cast the output tracking distribution of a deterministic algorithm, for example a correlation surface arising from template matching, as an approximate posterior distribution. The arguments we put forth for this consideration are as follows.
First, we choose the template matching scheme as a representative algorithm from the deterministic class of trackers. Secondly, in accordance to this scheme, we consider the sample index in eqn.(1) as a location index on the search space for template matching.
In the template based tracking scheme, the search for the template at any tracking step is centered on the inferred location of the template in the previous step. Hence, the terms in the weight proportionality in eqn.(2) can be indexed as follows.

$$w_k^i \propto \frac{p\left(\frac{z_k}{x_k^i}\right) p\left(\frac{x_k^i}{x_{k-1}}\right)}{q\left(\frac{x_k^i}{x_{k-1}, z_{1:k}}\right)} w_{k-1}^i, 1 \leq i \leq N \quad (3)$$

Notably, the index on the previous state x_{k-1} is removed. This operation resembles the resampling procedure used in SMC filtering. In this case, resampling the distribution at the instant $k-1$ results in representing the distribution by a single Kronecker-delta at the estimated location x_{k-1}. The reader must note that we use the same representation for both the estimated location and the random variable representing the state.
For the sake of tractability in SMC filtering, the importance sampling density in eqn.(3) is usually taken to be the Markovian prior as follows;

$$q\left(\frac{x_k^i}{x_{k-1}, z_{1:k}}\right) = p\left(\frac{x_k^i}{x_{k-1}}\right). \quad (4)$$

In the template tracking scheme, if we denote $S(x_{k-1})$ as the search area centered on the Kronecker-delta at the estimated location x_{k-1}, then we can define the Markovian prior as follows.

$$p\left(\frac{x_k^i}{x_{k-1}}\right) = \frac{1}{N} \Sigma_{x_k^i \in S(x_{k-1})} \delta\left(x_k - x_k^i\right), 1 \leq i \leq N \quad (5)$$

Bearing upon such a prior, the weights after resampling, are just proportional to the likelihood distribution,

$$w_k^i \propto p\left(\frac{z_k}{x_k^i}\right), 1 \leq i \leq N. \quad (6)$$

These weights, after normalization, result in the following probabilities;

$$\bar{w}_k^i = \frac{w_k^i}{\Sigma_{i=1}^N w_k^i}, 1 \leq i \leq N. \quad (7)$$

We then have the required approximation of the posterior distribution as;

$$p\left(\frac{x_{0:k}}{z_{1:k}}\right) \approx \Sigma_{i=1}^N \bar{w}_k^i \delta\left(x_{0:k} - x_{0:k}^i\right) \quad (8)$$

From eqn.(8) above we can derive an approximation for the instantaneous filtering distribution $p\left(\frac{x_k}{z_{1:k}}\right)$ at time k as follows;

$$p\left(\frac{x_k}{z_{1:k}}\right) = \int p\left(\frac{x_{0:k}}{z_{1:k}}\right) dx_{0:k-1} \approx \Sigma_{i=1}^N \bar{w}_k^i \delta\left(x_k - x_k^i\right) \quad (9)$$

Retracing the last few steps, we see that the likelihood distribution in the context of template based tracking is proportional to the unnormalized correlation surface (in the sense of a probability distribution). Therefore, we can consider the normalized correlation surface as an approximation to the filtering distribution of the tracked state. In the preceding arguments we have established a consistency between deterministic tracking and SMC filtering. The consequence is that, it makes it possible to treat both classes of algorithms uniformly for further analysis of their performance based on the differential entropy of the filtering distribution. In the following section we estimate the differential entropy of the tracked state.

3. DIFFERENTIAL ENTROPY OF THE TRACKED STATE

In the rest of the paper the terms entropy and differential entropy are used interchangeably.
The conditional entropy of random variable X_k conditioned on several random variables $\{Z_{1:k}\}$ is defined below for convenience [4].

$$h\left(\frac{x_k}{z_{1:k}}\right) = -\int p\left(z_{1:k}\right) \int p\left(\frac{x_k}{z_{1:k}}\right) \ln p\left(\frac{x_k}{z_{1:k}}\right) dx_k dz_{1:k} \quad (10)$$

In eqn.(10) we are required to evaluate the integral over the space of all possible data (all possible image sequences), which is not tractable. But, in practice the image sequence data is given sequentially and the filtering distribution is only inferred (this is similar to evaluating probability likelihoods, wherein the event has already occurred and the probability of a particular outcome is sought). Therefore, $p\left(z_{1:k}\right)$ degenerates to a Dirac-delta function at $z_{1:k}^s$, the given data sequence:

$$p\left(z_{1:k}\right) = \delta\left(z_{1:k} - z_{1:k}^s\right) \quad (11)$$

From this tractability assumption in eqn.(11) above we reduce the computation of the entropy to the following expression.

$$h\left(\frac{x_k}{z_{1:k}}\right) = -\int p\left(\frac{x_k}{z_{1:k}^s}\right) \ln p\left(\frac{x_k}{z_{1:k}^s}\right) dx_k \quad (12)$$

The starting point to approximate the entropy above is the weighted approximation of the filtering distribution in eqn.(9). We consider this weight distribution as an approximation of the *continuous filtering distribution*. We then construct approximations of the continuous marginal filtering distributions in each dimension (here two). For convenience we denote these marginals as $f_{x_{k1}}$ and $f_{x_{k2}}$.
For densities $f_{x_{k1}}$ and $f_{x_{k2}}$ we approximate their respective entropies $h_{x_{k1}}$ and $h_{x_{k2}}$ using the techniques presented in [3]. Making the worst case assumption that the variables in each dimension are independent, we arrive at a numerical approximation of the entropy in eqn.(12) as follows:

$$h\left(\frac{x_k}{z_{1:k}}\right) \approx h_{x_{k1}} + h_{x_{k2}}. \quad (13)$$

In the above steps we dealt with the differential entropy of the tracked state within a SMC approach. In the next section, we develop a full fledged tracker influenced by the time series behavior of its differential entropy.

4. RANDOMIZED FEATURE TRACKER

In the starting frame of a test color sequence, we mark the object to be tracked by a bounding box. Inside this bounding box, we choose a fixed number M coordinate locations using a *uniform random number* generator. Each of these coordinates form the centre of a feature. The *appearance model* (template) of each feature is an image patch of a preset size around its centre. We note that none of the templates are allowed to exceed the physical limits of the bounding box. Further, we associate each feature with an *object reference vector* connecting its centre to the centre of the bounding box (which is taken to be the object centre). Therefore, we have the set of templates and their corresponding reference vectors $\{T_j, r_j\}_{j=1}^{M}$. Given this initial configuration, the step wise tracking procedure described below is iterated.

1. **Approximating the continuous filtering distribution from M features**
   From time step $k-1$ to k, $k \geq 1$, we track each feature individually using a standard normalized cross-correlation scheme. The resulting set of correlation surfaces are denoted as $\{C_j\}_{j=1}^{M}$. The unnormalized weights in eqn.(6) are computed as follows:
   $$w_k^i \propto \Sigma_{j=1}^{M} \left[C_j + r_j\right](i) \tag{14}$$
   From eqn(14) and eqns.(7,9) we arrive at the required approximation. Using this approximation we compute the entropy of the filtering distribution (See Section 3) and record it in a time series $[h_1, h_2 \ldots h_k]$.
2. **Feature rejection and residual resampling**
   At tracking step $k : k \geq 1$, we perform the following test.
   If, $|h_k - h_1| \geq \tau$, where τ is a preset threshold.
   then, attempt to reject and resample features.
   else, retain the feature set.
   Rejection: The translation vectors (from step $k-1$ to k) of the features are clustered in the two dimensional motion space and outliers discarded. The filtering distribution is re-approximated using the correlation surfaces of the inliers.
   Resampling: The feature set is replenished by sampling *new feature locations* from the filtering distribution and associating templates (surrounding image patches) to each.
3. **State estimation**
   The state at step $k : x_k$ is estimated as the mean of the approximate filtering distribution. The reference vectors $\{r_j\}_{j=1}^{M}$ are then re-estimated given the state estimate.

To summarise the preceeding procedure we present a flow chart of the proposed tracking approach in Fig.(1).

In the next section we present experimental results of the proposed scheme alongwith discussions.

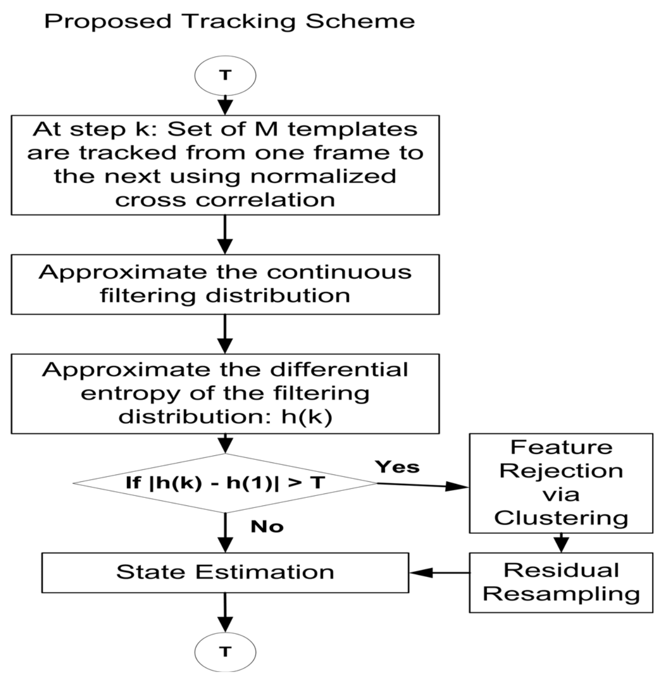


Fig. 1. Flow Chart of the Randomized Feature Tracker

(a). Frame 1 (b). Frame 18 (c). Frame 60 (d). Frame 78 (e). Frame 91 (f). Frame 102 (g). Frame 110 (h). Frame 115 (i). Frame 172 (j). Frame 173

Fig. 2. Randomized Feature Tracking on the Snakeeyes sequence.

5. EXPERIMENTS AND DISCUSSIONS

In all the experiments 8 features were used. The template size for the features were chosen to be 21×21 pixels and the search area for the features were 61×61 pixels. The threshold for rejection and resampling was empirically set at 0.1 nats (natural logarithm). The test video sequences were cinema sequences with human heads and aerial videos of moving vehicles from the PETS 2005 database [6]. We now present results of head tracking with the randomized feature-tracker and the corresponding time series recording of the entropy in Figs. 2 and 3 respectively. We have chosen the sample frames above to draw the readers attention towards the particular ability of the algorithm to handle out of the image plane motions and extreme illumination changes (See Figs. 2(d),2(e),2(f),2(g)). It is also

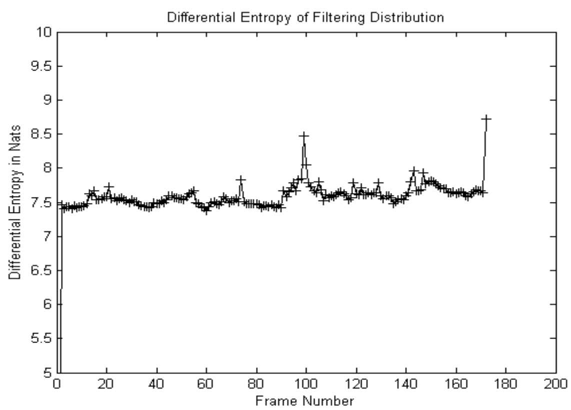
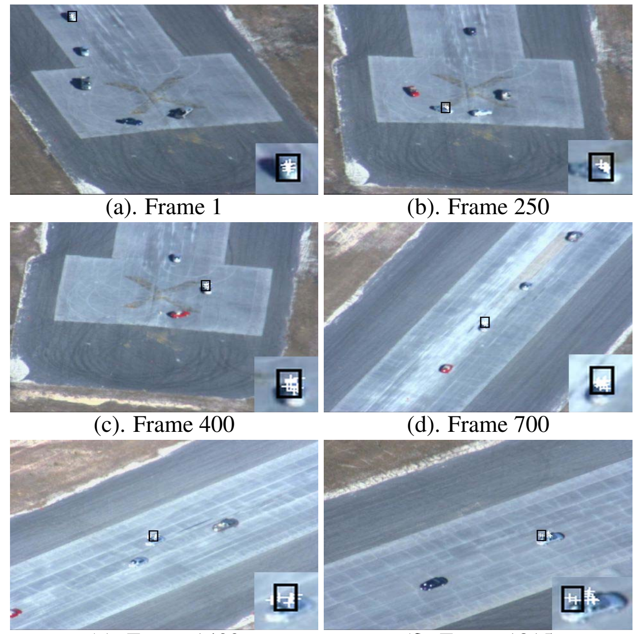


Fig. 3. A time series plot of the differential entropy of the tracked state

informative to observe the corresponding entropy increase at these phases(See Fig. 3).
At any point in the tracking sequence, each of the M features possibly have a different age than the rest: meaning, some features have their associated templates from the distant past, some relatively new and the remaining absolutely new. This mixture of templates is richly diverse and is a key ingredient in this tracking scheme. The novelty here is that, neither is it necessary to arrive at a single, complex, combined appearance model, for instance via a linear combination of appearance models from various instants in the tracking sequence nor perform template adaptation for each feature. Instead, the same or superior results can be obtained by a *self adaptive set of age variant appearance templates*.
Apart from the appearance adaptability of this scheme, two other differentiating factors exist in its favor. The first is the conspicuous absence of a feature selection or extraction scheme, as in current research [7]. Such selection algorithms are usually plagued by non-repeatability of selected features in successive frames. Further, these schemes only tend to qualify image patches with high gradient as the best features. But from these experiments, it is often seen that even visually smooth templates are good contenders for tracking. The second is that unlike feature tracking schemes using KLT like trackers [7], the numbers of features required are very few (usually not more than 10).
The primary limitations of this scheme are the absence of an occlusion handler (the occlusion though is clearly captured by a large increase in the entropy. See Fig. 3) and the presence of drift in the state estimate due to *aggregation* of the features over time. The drift aspect is highlighted in the result samples shown in Fig. 4.

6. CONCLUSIONS

Every characteristically different tracking scheme, say template based tracking or SMC filtering has its well known inherent limitation. It is expected that a mutually beneficial combination of these different schemes would result in robust tracking. In this large context of information fusion lies the objective of this contribution, which primarily was to bring forth a methodology or framework to study tracking performance. The framework relied upon casting tracking distributions as Monte Carlo approximations of filtering distributions. Once cast, it was then possible to assess different tracking strategies using a common measure. Towards this end, an information-theoretic measure of performance was chosen based on its natural properties and convenient usage approximations. The quality assessment measure was then put to test on a novel strategy of randomized feature tracking and conclusive results presented. The limitations of the randomized feature tracking are tracking drift, lack of scale adaptability and occlusion handling. However, the scheme makes for an ideal candidate for complementary tracking with, say, a probabilistic color based tracker.



(a). Frame 1 (b). Frame 250
(c). Frame 400 (d). Frame 700
(e). Frame 1400 (f). Frame 1815

Fig. 4. Randomized Feature Tracking on a PETS 2005 sequence.

7. REFERENCES

[1] Mathias Kolsch and Matthew Turk, "Fast 2d tracking with flocks of features and multi-cue integration," in *CVPR Workshop*, 2004. 1

[2] Kevin Nickels and Seth Hutchinson, "Estimating uncertainty in ssd-based feature tracking," *Image and Vision Computing*, vol. 20, pp. 47–68, 2002. 1

[3] Aapo Hyvarinen, "New approximations of differential entropy for independent component analysis and projection pursuit," Tech. Rep., Helsinki University of Technology, 1997. 1, 2

[4] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley and Sons, Inc., 1991. 1, 2

[5] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50(2), pp. 174–188, 2002. 2

[6] Robert T. Collins, Xuhui Zhou, and Seng Keat Teh, "An open source tracking testbed and evaluation web site," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2005. 3

[7] J.Shi and C.Tomasi, "Good features to track," in *CVPR, Seattle*, 1994. 4