

BAYESIAN VISUAL TRACKING WITH EXISTENCE PROCESS

J. Vermaak⁽¹⁾, *S. Maskell*⁽²⁾, *M. Briers*⁽²⁾, *P. Pérez*⁽³⁾

(1) Cambridge Univ. Eng. Dpt, Cambridge, U.K., jv211@eng.cam.ac.uk

(2) QinetiQ Ltd., Malvern, U.K., {s.maskell,m.briers}@signal.qinetiq.com

(3) IRISA/INRIA Rennes, France, perez@irisa.fr

ABSTRACT

Most object tracking approaches either assume that the number of objects is constant, or that information about object existence is provided by some external source. Here, we show how object existence can be rigorously integrated within the Bayesian single and multiple object tracking framework. We provide a general treatment that impacts as little as possible on existing tracking algorithms, so that software can be reused, and that allows implementation with Kalman Filters, Extended Kalman Filters, Particle Filters, *etc.* We apply the proposed framework to colour-based tracking of multiple objects.

1. INTRODUCTION AND ASSUMPTIONS

Visual tracking involves the detection and recursive localisation of objects within video frames. Most tracking algorithms assume the number of objects to be constant in the sequence. If information about object appearance and disappearance is available from some external source, the algorithms are sometimes deterministically reconfigured for the updated number of objects. However, the birth and death of objects are themselves uncertain events. Object existence should thus be treated jointly with other unknowns within a probabilistic framework that can account for all the expected ambiguities. One way to achieve this is to associate with each object a binary existence variable indicating whether the object is active or not. A suitable prior process can then be specified for the existence variables, facilitating joint inference with the state and data association processes.

This idea is not new, and have been considered before [1, 2, 3]. In this paper, however, we attempt to establish a general framework that is independent of a specific implementation strategy, so that existing tracking architectures can be re-used. We only assume that, at each instant, a sparse set of measurements is obtained by an appropriate analysis of the raw image. One unknown of the tracking system is then the association of these measurements to track(s) or to clutter. In this context, our approach is based on the (Joint) Probabilistic Data Association filter [4].

Modelling assumptions. For single object tracking we will be interested in recursively estimating the objec-

t state \mathbf{x}_t and existence variable $e_t \in \{0, 1\}$, based on a sequence of observations $\mathbf{y}^t = (\mathbf{y}_1 \cdots \mathbf{y}_t)$. At time t observations $\mathbf{y}_t = (\mathbf{y}_t^1 \cdots \mathbf{y}_t^{M_t})$ comprise M_t detections, which are unlabelled and may be due to the object or clutter. Assuming that at most one measurement at a time can be due to the object, we introduce the association variable $a_t \in \{0, 1 \cdots M_t\}$, where $a_t = j$, $j > 0$ if the j -th observation is due to the object, and $a_t = 0$ if the object is undetected. The complete set of unknowns at time t is thus $\mathcal{X}_t = \{\mathbf{x}_t, e_t, a_t\}$, for which we assume a Markovian prior

$$p(\mathcal{X}_t | \mathcal{X}_{t-1}) = p(a_t | e_t) p(\mathbf{x}_t | \mathbf{x}_{t-1}, e_t, e_{t-1}) p(e_t | e_{t-1}).$$

We assume the existence variable to follow a discrete Markov chain, parameterised by the object birth and death probabilities, *i.e.* $P_b = p(E_t | \bar{E}_{t-1})$ and $P_d = p(\bar{E}_t | E_{t-1})$, where E_t and \bar{E}_t stand for $e_t = 1$ and $e_t = 0$, respectively. Conditional on the existence variables the Markov process for the object state is specified by

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, e_t, e_{t-1}) = \begin{cases} \delta_{\mathbf{u}}(\mathbf{x}_t) & \text{if } \bar{E}_t \\ p_0(\mathbf{x}_t) & \text{if } (E_t, \bar{E}_{t-1}) \\ p_D(\mathbf{x}_t | \mathbf{x}_{t-1}) & \text{if } (E_t, E_{t-1}), \end{cases}$$

where \mathbf{u} is the consuming object state that corresponds to the object not existing, p_0 is the initial state distribution, and p_D is the object dynamic model. The association prior is independent of previous temporal information. Finally, we will need two models, $p_T(\mathbf{y}_t^* | \mathbf{x}_t)$ and $p_C(\mathbf{y}_t^*)$, for the measurements originating from the object and clutter, respectively. Conditional on the state, existence and association variables, we assume the measurements to be independent.

2. BAYESIAN TRACKING WITH EXISTENCE

For single object tracking we are interested in recursively estimating the joint posterior

$$p(\mathbf{x}_t, E_t | \mathbf{y}^t) = p(E_t | \mathbf{y}^t) \sum_{a_t} p(a_t | E_t, \mathbf{y}^t) p(\mathbf{x}_t | E_t, a_t, \mathbf{y}^t).$$

For brevity we will use the notation $P_{E_t} = p(E_t | \mathbf{y}^t)$. Using conditional independence of measurements

$$p(\mathbf{x}_t | E_t, a_t, \mathbf{y}^t) = \frac{p_T(\mathbf{y}_t^{a_t} | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1})}{p(\mathbf{y}_t^{a_t} | E_t, \mathbf{y}^{t-1})}, \quad (1)$$

where the denominator

$$p(\mathbf{y}_t^* | E_t, \mathbf{y}^{t-1}) = \int p_T(\mathbf{y}_t^* | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1}) d\mathbf{x}_t \quad (2)$$

is the measurement prediction distribution. Finally,

$$p(\mathbf{x}_t | E_t, \mathbf{y}^t) = \sum_{j=0}^{M_t} \beta_j \frac{p_T(\mathbf{y}_t^j | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1})}{p(\mathbf{y}_t^j | E_t, \mathbf{y}^{t-1})}, \quad (3)$$

where $\beta_j = p(a_t = j | E_t, \mathbf{y}^t)$, and, with a slight abuse of notation, $p_T(\mathbf{y}_t^0 | \mathbf{x}_t) = 1$.

Both expressions (2) and (3) require the computation of the state prediction distribution $p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1})$, which can be expressed as

$$p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1}) = \sum_{e_{t-1}} p(e_{t-1} | E_t, \mathbf{y}^{t-1}) \times \int p(\mathbf{x}_t | \mathbf{x}_{t-1}, E_t, e_{t-1}) p(\mathbf{x}_{t-1} | e_{t-1}, \mathbf{y}^{t-1}) d\mathbf{x}_{t-1}.$$

It is essentially a mixture of prediction distributions, with

$$p(e_{t-1} | E_t, \mathbf{y}^{t-1}) = \frac{p(E_t | e_{t-1}) p(e_{t-1} | \mathbf{y}^{t-1})}{\sum_{f_{t-1}} p(E_t | f_{t-1}) p(f_{t-1} | \mathbf{y}^{t-1})}.$$

The state prediction distribution finally follows as

$$p(\mathbf{x}_t | E_t, \mathbf{y}^{t-1}) = \frac{P_b(1-P_{E_{t-1}})p_0(\mathbf{x}_t) + (1-P_d)P_{E_{t-1}}p_{\text{pred}}(\mathbf{x}_t | \mathbf{y}^{t-1})}{P_b + P_{E_{t-1}}(1-P_b - P_d)},$$

with $p_{\text{pred}}(\mathbf{x}_t | \mathbf{y}^{t-1}) = \int_{\mathbf{x}_{t-1}} p_D(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | E_{t-1}, \mathbf{y}^{t-1})$.

Note that, if previous state distribution $p(\mathbf{x}_{t-1} | E_{t-1}, \mathbf{y}^{t-1})$ is a mixture with N_{t-1} components, the new one is a mixture with $(N_{t-1} + 1)(M_t + 1)$ components. In case of mixture modelling, a mixture reduction technique is thus required to avoid exponential growth in the number of components.

We still need the marginal posterior of the existence variable $p(E_t | \mathbf{y}^t)$ and the conditional posterior of the association variable $\beta_j = p(a_t = j | E_t, \mathbf{y}^t)$ to update the filtering distribution. This is easily done by summation once the joint posterior $p(e_t, a_t | \mathbf{y}^t)$ is known. In what follows we will show how this joint can be computed.

The joint posterior for the existence and association variables can be expressed as

$$p(e_t, a_t | \mathbf{y}^t) = \int p(\mathbf{x}_t, e_t, a_t | \mathbf{y}^t) d\mathbf{x}_t \times p(a_t | e_t) p(e_t | \mathbf{y}^{t-1}) p_C(\mathbf{y}_t^{R(e_t, a_t)}) p(\mathbf{y}_t^{a_t} | e_t, \mathbf{y}^{t-1}), \quad (4)$$

where $R(e_t, a_t)$ is the set of indices of clutter measurements under the existence-association hypothesis (e_t, a_t) . These measurements are all evaluated under the clutter model. In the above $p(a_t | e_t)$ is the association prior, $p(e_t | \mathbf{y}^{t-1})$ is the prediction for the existence variable, $p_C(\mathbf{y}_t^{R(e_t, a_t)})$ is the likelihood for the clutter measurements, and $p(\mathbf{y}_t^{a_t} | e_t, \mathbf{y}^{t-1})$ is the measurement prediction. Note that the measurement prediction term disappears for non-existing objects and

missed detections, *i.e.* $e_t = 0$ and $a_t = 0$. Thus, the only unknown is the prediction for the existence variable

$$p(e_t | \mathbf{y}^{t-1}) = \sum_{e_{t-1}} p(e_t | e_{t-1}) p(e_{t-1} | \mathbf{y}^{t-1}). \quad (5)$$

Substituting from the modelling equations this becomes

$$p(\bar{E}_t | \mathbf{y}^{t-1}) = (1 - P_b)(1 - P_{E_{t-1}}) + P_d P_{E_{t-1}} \\ p(E_t | \mathbf{y}^{t-1}) = P_b(1 - P_{E_{t-1}}) + (1 - P_d)P_{E_{t-1}}.$$

In practice the expression (4) must be evaluated for all the valid existence-association pairs. The probabilities can then be normalised by dividing each with the sum over all the valid pairs. For \bar{E}_t there is only one possible association, *i.e.* $a_t = 0$, so that all the observations are due to clutter. For E_t and M_t measurements there are $M_t + 1$ possible associations, *i.e.* $a_t = j$, $j = 0 \dots M_t$, with $a_t = 0$ if the object is undetected, and $a_t = j > 0$ if the j -th measurement originated from the object.

The framework above is entirely general, and can be implemented in many different ways. If, for example, the filtering distribution $p(\mathbf{x}_t | E_t, \mathbf{y}^t)$ is to be represented by a Gaussian mixture, the mixture components can be obtained by the Kalman filter for linear Gaussian models, and by the extended or unscented Kalman filters for non-linear and/or non-Gaussian models. For models of the latter kind it may sometimes be beneficial to adopt a particle representation, and use sequential importance sampling techniques to update the filtering distribution.

Multi-object tracking. For multi-object tracking the distribution of interest is the joint posterior of all the object states and existence variables, *i.e.* $p(\mathbf{x}_t^{1:K}, E_t^{1:K} | \mathbf{y}^t)$, where K is the maximum number of objects that can be tracked and the superscript denotes the range over all the objects. In a sense K is the number of ‘‘slots’’ available for tracked objects, *i.e.* when an object is killed its slot may be arbitrarily reassigned to a newly detected future object.

To avoid exploring the joint state-space for all the objects, which increases exponentially in complexity with the number of objects, we will assume, like in the JPDA [4], that the joint posterior factorises over the objects, *i.e.*

$$p(\mathbf{x}_t^{1:K}, E_t^{1:K} | \mathbf{y}^t) = \prod_{k=1}^K p(\mathbf{x}_t^k | E_t^k, \mathbf{y}^t) p(E_t^k | \mathbf{y}^t), \quad (6)$$

and it will be our objective to maintain these marginal distributions. For each individual object the state filtering proceeds in exactly the same manner as with single object. However, the computation of the existence and association probabilities cannot be done independently for each of the objects. Since the origins of the measurements are unknown the existence and association variables for the individual objects are coupled, and have to be treated jointly. Thus, we

need to consider the joint posterior $p(e_t^{1:K}, a_t^{1:K} | \mathbf{y}^t)$ of the existence and association variables for all the objects. The required marginals and conditionals for the individual objects are easily obtained by summation.

Using computations similar to those leading to (4) the required joint is easily shown to be

$$p(e_t^{1:K}, a_t^{1:K} | \mathbf{y}^t) \propto p(M_t^C) p_C(\mathbf{y}_t^{R(e_t^{1:K}, a_t^{1:K})}) \\ \times \prod_{k=1}^K \left[p(a_t^k | a_t^{1:k-1}, e_t^k) p(e_t^k | \mathbf{y}^{t-1}) p(\mathbf{y}_t^{a_t^k} | e_t^k, \mathbf{y}^{t-1}) \right],$$

where we have made use of the factorised form of the association prior (see [5] for details):

$$p(a_t^{1:K} | e_t^{1:K}) = p(M_t^C) \prod_{k=1}^K p(a_t^k | a_t^{1:k-1}, e_t^k), \quad (7)$$

and $p(M_t^C)$ denotes the Poisson prior over the number of clutter measurements.

Again it is necessary to enumerate all the valid joint existence-association hypotheses. Whereas this is tractable for a single object, the computational complexity for a direct enumeration of the joint hypotheses in the multi-object case grows super-exponentially in the number of objects. Measurement *gating* is a crucial first step to reduce the number of hypotheses. But for a large number of objects and measurements gating alone is not sufficient. Additional speed up can be obtained by efficient hypothesis management (EHM)[6] or Monte Carlo approximation [5].

3. APPLICATION

In this section we demonstrate the performance of the proposed framework for colour-based tracking of multiple objects. As proposed in [7], a reference colour histogram is chosen at initialisation time to characterise the global appearance of the objects of interest. This type of description has proved useful for robust tracking of objects despite dramatic changes of shapes and poses. The ‘‘mean-shift’’ (MS) technique of Comaniciu *et al.* allows the tracking of a single object by deterministic and iterative search of the best location in the current frame. We turn this technique into a detection module that produces the measurements handled by our probabilistic framework. More precisely, MS is run from a set of starting positions that are evenly spread over the image plane, each one thus providing a candidate location and window size. We only retain those with associated colour histograms sufficiently close to the reference one (Bhattacharyya coefficient larger than 0.6). These constitute the M_t measurements at time t .

We define the state of each object as its 2D location and scale in the image, along with the velocities of these quantities. Thus, the relationship between the object state and the detection is linear. For the object dynamics we adopt a near constant velocity model that is independent for each of the

state components, and we assume the measurements to be corrupted by Gaussian noise that is independent for each of the measurement components. This implies that the state filtering distribution for each of the objects can be represented by a Gaussian mixture, and that the update to compute the new mixture components can be performed by the Kalman filter.

We initialise the algorithm with no active objects. For each frame new objects are initialised on all measurements that are not associated with any of the existing objects. The initial state distribution is set to be a Gaussian centred on the measurement and zero velocity, and the initial existence probability is set to 0.1. New objects are only verified to be true objects once their corresponding existence probability exceeds 0.9. Thus, if the measurement was due to clutter it is unlikely to be reinforced over time, and the existence probability will not increase. True object measurements, however, will tend to be persistent over time, which is soon reflected by the existence probability increasing to a value close to unity. When objects leave the scene measurements disappear, and the existence probability decreases. Once this probability falls below 0.1 we deem the object to have left the scene, and deactivate the corresponding tracker. For active objects we perform mixture reduction using the strategy described in [8], and we take object estimates to be the MAP value of the resulting Gaussian mixture.

As a experimental example, we address the problem of tracking football players in video sequences 1. The objective here is to track all the football players in white. It is clear that the number of players in view changes throughout the sequence. Note also that the detector is not perfect, *i.e.* false positives (clutter measurements) and missed detections are plentiful.

Despite the presence of clutter measurements and missed detections the algorithm is able to successfully detect and track all the football players for as long as they are in view of the camera system (Fig. 1 and videos available at www.sigproc.eng.cam.ac.uk/~jv211/videos/exist). Even though objects are wrongly initialised on clutter measurements they are never verified by the algorithm, and the corresponding trackers are deactivated after a few time steps (Figs. 1-2). The algorithm is able to smooth the location and scale estimates provided by the detector, and to bridge over missed detections without losing track.

4. CONCLUSIONS

We have presented a general framework for the joint treatment of object tracking and existence. The framework is able to robustly deal with an unknown and variable number of objects by efficiently computing the posterior distribution of the object existence and association variables. We have applied the algorithm to colour-based detection and tracking of multiple objects in videos, with a mixture Kalman

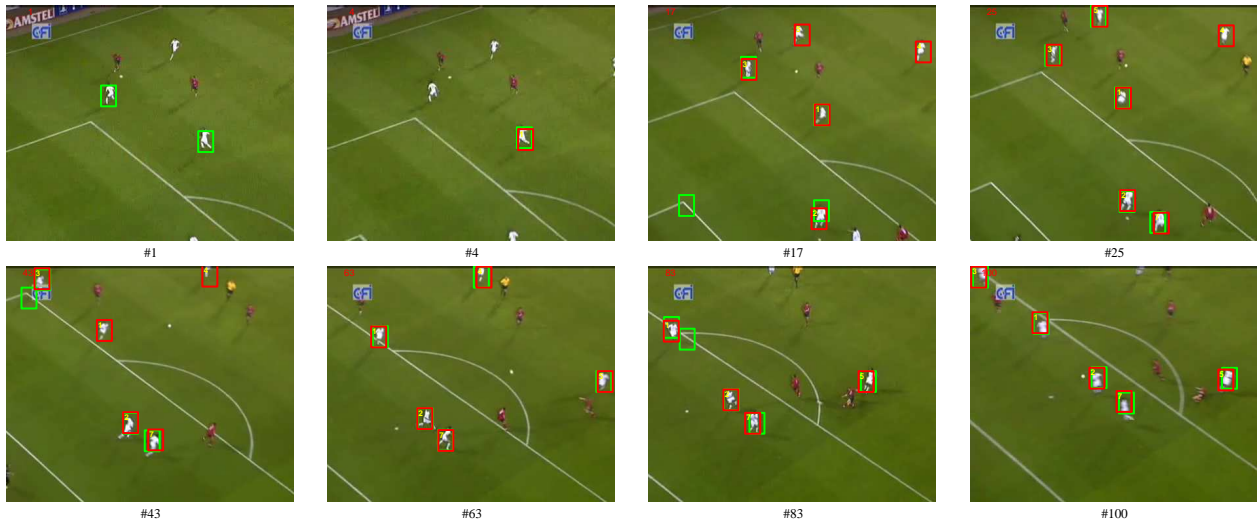


Fig. 1. Detection and tracking results. Detections (green boxes) include false positives and missed detections. Nonetheless, the algorithm successfully detects and tracks all visible white players (red boxes), whose number changes continuously.

filter implementation of the general framework. Despite a constantly changing number of objects and a large number of missed detections and false positives the algorithm was able to successfully detect and track all the objects (football players in the example) for as long as they are within the field of view. Future work will focus on implementations for other visual tracking tasks and extensions to cope with a very large number of objects.

5. REFERENCES

- [1] D. Mušicki, R. Evans, and S. Stanković, “Integrated probabilistic data association,” *IEEE Trans. Automatic Control*, vol. 39, no. 6, pp. 1237–1241, 1994.
- [2] D. Mušicki and R. Evans, “Joint integrated probabilistic data association – JIPDA,” in *Proc. Intl. Conf. Inf. Fusion*, 2002, vol. 2, pp. 1120–1125.
- [3] M. Rollason and Salmond. D., “A particle filter for track-before-detect of a target with unknown amplitude,” in *IEE Target Tracking: Algo. and Appl.*, 2001, vol. 1, pp. 14/1–14/4.
- [4] T. Kirubarajan and Y. Bar-Shalom, “Probabilistic data association techniques for target tracking in clutter,” *Proc. IEEE*, vol. 92, no. 3, pp. 536–557, 2004.
- [5] J. Vermaak, S.J. Godsill, and P. Pérez, “Monte Carlo filtering for multi-target tracking and data association,” *IEEE Trans. Aerospace and Elect. Systems*, 2005, to appear.
- [6] S. Maskell, M. Briers, and R. Wright, “Fast mutual exclusion,” in *Proc. SPIE*, 2004.
- [7] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift,” in *Proc. Conf. Comp. Vision and Pattern Recog.*, 2000, pp. II: 142–149.
- [8] D.J. Salmond, “Mixture reduction algorithms for target tracking in clutter,” in *Signal and Data Processing of Small Targets*, *SPIE 1305*, 1990, pp. 434–445.

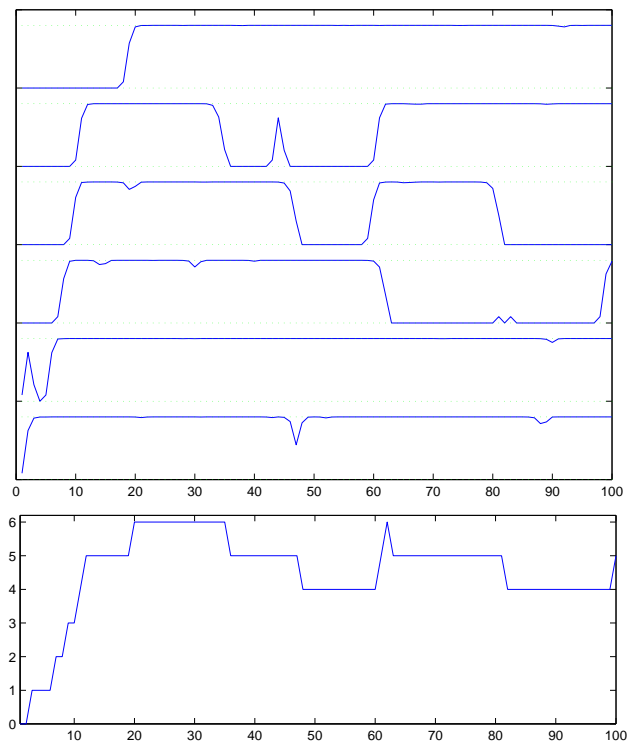


Fig. 2. Existence probabilities and number of objects. (Top) The existence probabilities for true objects soon ramp up to unity, whereas those for wrongly initialised objects fall to zero. (Bottom) The estimated number of active objects accurately reflects the information in the sequence.