

A Unifying Framework for Multi-Target Tracking and Existence

Jaco Vermaak
Cambridge University Engineering Department
Cambridge, U.K.
jv211@eng.cam.ac.uk

Simon Maskell, Mark Briers
QinetiQ Ltd., Malvern Technology Centre
Malvern, U.K.
{s.maskell,m.briers}@signal.qinetiq.com

Patrick Pérez
IRISA / INRIA Rennes, France, perez@irisa.fr

Abstract – *Most target tracking approaches either assume that the number of targets is constant throughout the time horizon of interest, or that information about target existence (birth and death) is provided by some external source. Here we show how target existence can be integrated within the tracking framework in a rigorous way. The notion of existence is not new, and has been considered before in e.g. [1, 2]. We provide here a general probabilistic treatment that impacts as little as possible on existing tracking algorithms so that legacy tracking software (and more generally target tracking architectures) can be reused. We first show how the notion of existence can be incorporated into a single target tracking framework (retaining algorithmic invariance). To place the probabilistic recursions into context we relate this single target tracking architecture to the Probabilistic Data Association [3, 4] filter. We then extend the single target results to incorporate existence for multi-target tracking and relate this to an importance sampling implementation of the Joint Probabilistic Data Association (JPDA) [5, 3, 6] framework. The treatment presented is entirely general and so facilitates implementation with Kalman Filters, Extended/Unscented Kalman Filters, Particle Filters, etc, i.e. the approach developed is invariant to the filtering and data association mechanisms used, and therein lies the novelty. We apply the proposed framework to the difficult problem of tracking football players in video sequences, where we adopt a mixture Kalman filter implementation.*

Keywords: Multi-target tracking, existence, data association, mixture Kalman filter.

1 Introduction

Multi-target tracking involves the detection and recursive localisation of targets within the observation range of a sensing system. Many popular tracking algorithms (such as Global Nearest Neighbour, JPDA, etc.) focus on the state filtering and data association aspects of the problem,

and assume the number of targets to be constant over the time horizon of interest. If information about target appearance and disappearance is available from some external source, the algorithms are sometimes deterministically reconfigured for the updated number of targets.

However, the birth and death of targets are themselves uncertain events, and a deterministic reconfiguration of the algorithms is likely to lead to errors in ambiguous situations. Ideally target existence should be treated jointly with the target state and data association processes within a probabilistic framework that can account for all the expected ambiguities. To enable the deployment of the vast body of existing multi-target tracking and data association algorithms such a framework should impact as little as possible on existing tracking architectures.

One way to achieve this is to associate with each target a binary existence variable indicating whether the target 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. The existence idea is not new, and has been considered before in the context of single [1] and multi-target tracking [2]. More recently it has also been applied to the track-before-detect problem using particle filters [7]¹.

In this paper we attempt to establish a general unifying framework for multi-target tracking and existence that is independent of a specific implementation strategy. We first consider the single target case, and for reference show how existence can be combined with PDA based data association. We then extend the results to the multi-target setting and briefly discuss an importance sampling implementation of the framework within a JPDA setting. Both the single and multi-target scenarios require enumeration of

¹Whilst every effort has been made to provide a complete set of references, it is the authors' opinion that the framework provided here allows all related publications to be unified; due to space constraints it has not been possible to explicitly link our work to the plethora of excellent publications on this subject.

all the joint existence-association hypotheses. This quickly becomes computationally infeasible for large numbers of targets and measurements. To mitigate these problems we consider a number of strategies to curb the exponential growth and keep the algorithms practically feasible.

The proposed framework makes no assumption about the implementation strategy. If the target state posteriors are to be represented by Gaussian mixtures the individual mixture components can be obtained by the Kalman filter for linear Gaussian models, and by the extended [3] or unscented [8] Kalman filters for models that are non-linear and/or non-Gaussian. For models of the latter kind it is often more effective to adopt a Monte Carlo representation for the state posterior, and use sequential importance sampling techniques [9] to perform the state filtering. This, also, is feasible within the proposed framework. It is the authors' opinion that this generic exposition forms the novelty within the paper, thus unifying the publications made in the literature to date.

The remainder of the paper is organised as follows: we outline the modelling assumptions in Section 2. Section 3 then shows how existence can be incorporated into a tracking framework for a single target, followed by the extension to multi-target tracking (within the JPDA) framework in Section 4. In Section 5 we apply the proposed framework to the difficult problem of tracking football players in video sequences. We adopt a Gaussian mixture representation for the target state, the components of which can be updated by the Kalman filter. Finally, we conclude with some remarks in Section 6.

2 Modelling Assumptions

For single target tracking we will be interested in recursively estimating the target 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 each time t the observations comprise M_t detections, which we will denote by $\mathbf{y}_t = (\mathbf{y}_t^1, \cdots, \mathbf{y}_t^{M_t})$. These detections are unlabelled and may be due to the target or clutter. Only one measurement at a time can be due to the target, but the target may also go undetected. To represent this uncertainty we introduce a target to measurement 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 target, and $a_t = 0$ if the target is undetected. The complete set of unknowns at time t is thus $\{\mathbf{x}_t, e_t, a_t\}$, for which we assume a Markovian prior of the form

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

We assume the existence variable to follow a discrete Markov chain, parameterised by the target birth and death

probabilities, *i.e.*

$$P_b = p(E_t | \bar{E}_{t-1}) \quad (2)$$

$$P_d = p(\bar{E}_t | E_{t-1}), \quad (3)$$

where we have used the abbreviations E_t and \bar{E}_t to mean $e_t = 1$ and $e_t = 0$, respectively. Conditional on the existence variables the Markov process for the target state is specified by

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, e_t, e_{t-1}) = \begin{cases} \delta_{\mathbf{u}}(\mathbf{x}_t) & \text{if } \bar{E}_t \\ p_0(\mathbf{x}_t) & \text{if } (E_t, \bar{E}_{t-1}) \\ p_D(\mathbf{x}_t | \mathbf{x}_{t-1}) & \text{if } (E_t, E_{t-1}), \end{cases} \quad (4)$$

where \mathbf{u} is the consuming target state that corresponds to the target not existing, p_0 is the initial state distribution, and p_D is the target dynamic model. The association prior is independent of previous temporal information. We will discuss it in more detail shortly.

We will need two models for the measurements: one for measurements originating from the target, and one for those from clutter. We will denote these by $p_T(\mathbf{y}_t^* | \mathbf{x}_t)$ and $p_C(\mathbf{y}_t)$, respectively. Conditional on the state, existence and association variables we will assume the measurements to be independent.

In the case of multi-target tracking with a maximum of K non-interacting targets the priors for the target state and existence variables factorise over the individual targets. We will denote the prior for the joint associations by $p(a_t^{1:K} | e_t^{1:K})$, where the superscript denotes the range over all the targets. We will make the common assumption that this distribution is the product of a uniform distribution for the association variables $a_t^{1:K}$, a binomial distribution for the number of target detections M_t^T , and a Poisson distribution for the number of clutter measurements M_t^C . (Note that such probabilistic assumptions can be relaxed; they are chosen here to provide the general reader some familiarity of the exposition.) For brevity we will continue to use the compact notation, as above, and only explicitly introduce the number of target detections and clutter measurements where the context requires it. As pointed out in [6], it is possible to partially factorise the association prior over the targets, where the components in the factorisation account for both the uniform prior over the valid associations and the binomial prior over the number of target detections. Thus, the association prior can be written as

$$p(a_t^{1:K} | e_t^{1:K}) = p(M_t^C) \prod_{k=1}^K p(a_t^k | a_t^{1:k-1}, e_t^k), \quad (5)$$

where $p(M_t^C)$ denotes the Poisson prior over the number of clutter measurements. Note that the ordering of the targets in the factorisation is arbitrary, and may be chosen randomly. We refer the reader to the reference above for the specific forms of the association prior components.

3 Single Target Tracking and Existence

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

$$p(\mathbf{x}_t, E_t | \mathbf{Y}^t) = p(\mathbf{x}_t | E_t, \mathbf{Y}^t) p(E_t | \mathbf{Y}^t). \quad (6)$$

For brevity we will use the notation $P_{E_t} = p(E_t | \mathbf{Y}^t)$. In what follows we will first show how to compute the conditional state posterior $p(\mathbf{x}_t | E_t, \mathbf{Y}^t)$. We will then show how to compute the joint posterior of the existence and association variables $p(e_t, a_t | \mathbf{Y}^t)$. From this joint probability measure the required marginals and conditionals are easily obtained by summation.

3.1 State Filtering

By introducing the association variable and using the law of total probability the target state posterior, also known as the filtering distribution, can be expressed as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^t) = \sum_{a_t} p(a_t | E_t, \mathbf{Y}^t) p(\mathbf{x}_t | E_t, a_t, \mathbf{Y}^t), \quad (7)$$

where $p(a_t | E_t, \mathbf{Y}^t)$ is the posterior association probability, and $p(\mathbf{x}_t | E_t, a_t, \mathbf{Y}^t)$ is the filtering distribution conditional on the association. Using Bayes' rule and the fact that the measurements at a scan are independent conditional on the target state, existence and association variables, the latter distribution can be obtained as

$$p(\mathbf{x}_t | E_t, a_t, \mathbf{Y}^t) = \frac{p_T(\mathbf{y}_t^{a_t} | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1})}{p(\mathbf{y}_t^{a_t} | E_t, \mathbf{Y}^{t-1})}, \quad (8)$$

where the denominator

$$p(\mathbf{y}_t^* | E_t, \mathbf{Y}^{t-1}) = \int_{\mathbf{x}_t} p_T(\mathbf{y}_t^* | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1}) d\mathbf{x}_t \quad (9)$$

is the measurement prediction distribution. By substituting this result into (7) the filtering distribution can be expressed as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^t) = \sum_{j=0}^{M_t} \beta_j \frac{p_T(\mathbf{y}_t^j | \mathbf{x}_t) p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1})}{p(\mathbf{y}_t^j | E_t, \mathbf{Y}^{t-1})}, \quad (10)$$

where we have introduced the notation $\beta_j = p(a_t = j | E_t, \mathbf{Y}^t)$. Note that $j = 0$ implies that the target is unassigned, and with a slight abuse of notation we will take the corresponding likelihood to be $p_T(\mathbf{y}_t^0 | \mathbf{x}_t) \propto 1$.

The expressions for the filtering distribution in (10) and the measurement prediction in (9) require the computation of the state prediction distribution $p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1})$. By introducing the previous state and existence variables, and using the rule of total probability, the state prediction can be

expressed as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1}) = \sum_{e_{t-1}} p(e_{t-1} | E_t, \mathbf{Y}^{t-1}) \times \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t | \mathbf{x}_{t-1}, E_t, e_{t-1}) p(\mathbf{x}_{t-1} | e_{t-1}, \mathbf{Y}^{t-1}) d\mathbf{x}_{t-1}. \quad (11)$$

It is essentially a mixture of prediction distributions, where the mixing terms follow from Bayes' rule as

$$p(e_{t-1} | E_t, \mathbf{Y}^{t-1}) = \frac{p(E_t | e_{t-1}) p(e_{t-1} | \mathbf{Y}^{t-1})}{\sum_{f_{t-1}} p(E_t | f_{t-1}) p(f_{t-1} | \mathbf{Y}^{t-1})}, \quad (12)$$

where we make the assumption that $p(E_t | e_{t-1}, \mathbf{Y}^{t-1}) = p(E_t | e_{t-1})$. By substituting from the prior description in Section 2 the state prediction distribution finally follows as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1}) = \frac{1}{P_b + P_{E_{t-1}}(1 - P_b - P_d)} \times \left[P_b(1 - P_{E_{t-1}}) p_0(\mathbf{x}_t) + (1 - P_d) P_{E_{t-1}} p_{\text{pred}}(\mathbf{x}_t | \mathbf{Y}^{t-1}) \right], \quad (13)$$

with

$$p_{\text{pred}}(\mathbf{x}_t | \mathbf{Y}^{t-1}) = \int_{\mathbf{x}_{t-1}} p_D(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | E_{t-1}, \mathbf{Y}^{t-1}) d\mathbf{x}_{t-1}. \quad (14)$$

Note that both the state and measurement prediction distributions are mixtures, with one component for each of the possible previous values of the existence variable. If the previous state distribution $p(\mathbf{x}_{t-1} | E_{t-1}, \mathbf{Y}^{t-1})$ is also a mixture, the total number of mixture components in the state and measurement distributions will exceed that in $p(\mathbf{x}_{t-1} | E_{t-1}, \mathbf{Y}^{t-1})$ by one. Thus, generally we can rewrite the state and measurement prediction distributions as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1}) = \sum_{i=1}^N \alpha_i p_i(\mathbf{x}_t | E_t, \mathbf{Y}^{t-1}) \quad (15)$$

$$p(\mathbf{y}_t^* | E_t, \mathbf{Y}^{t-1}) = \sum_{i=1}^N \alpha_i p_i(\mathbf{y}_t^* | E_t, \mathbf{Y}^{t-1}), \quad (16)$$

where α_i denotes the (generically defined) mixture component weights, which result from the propagation (and thus inclusion) of mixture components over time.

Substituting these expressions into (10) and rearranging terms the filtering distribution can be finally obtained as

$$p(\mathbf{x}_t | E_t, \mathbf{Y}^t) = \sum_{i=1}^N \sum_{j=0}^{M_t} \gamma_{ij} p_{ij}(\mathbf{x}_t | E_t, \mathbf{Y}^t), \quad (17)$$

with

$$p_{ij}(\mathbf{x}_t|E_t, \mathbf{Y}^t) = \frac{p_T(\mathbf{y}_t^j|\mathbf{x}_t)p_i(\mathbf{x}_t|E_t, \mathbf{Y}^{t-1})}{p_i(\mathbf{y}_t^j|E_t, \mathbf{Y}^{t-1})} \quad (18)$$

$$\gamma_{ij} = \frac{\alpha_i \beta_j p_i(\mathbf{y}_t^j|E_t, \mathbf{Y}^{t-1})}{\sum_{m=1}^N \alpha_m p_m(\mathbf{y}_t^j|E_t, \mathbf{Y}^{t-1})}. \quad (19)$$

Note that the filtering distribution is a mixture of $N(M_t+1)$ components, one for each possible combination of a measurement with a component of the state and measurement prediction distributions. Over time the number of components in this mixture will tend to grow exponentially, so that mixture reduction techniques are required. One such strategy for Gaussian mixtures is described in [10].

3.2 Existence and Association Posterior

From the discussion so far it is clear that we need the marginal posterior of the existence variable $p(E_t|\mathbf{Y}^t)$, and the conditional posterior of the association variable $\beta_j = p(a_t = j|E_t, \mathbf{Y}^t)$, $j = 0, 1 \dots M_t$, to update the filtering distribution. These are both features of the joint existence and association posterior $p(e_t, a_t|\mathbf{Y}^t)$, and are easily obtained by summation once the joint is known. In what follows we will show how this joint can be computed.

By introducing the target state and using Bayes' rule the joint posterior for the existence and association variables can be expressed as

$$\begin{aligned} p(e_t, a_t|\mathbf{Y}^t) &= \int_{\mathbf{x}_t} p(\mathbf{x}_t, e_t, a_t|\mathbf{Y}^t) d\mathbf{x}_t \\ &\propto p(a_t|e_t)p(e_t|\mathbf{Y}^{t-1})p_C(\mathbf{y}_t^{R(e_t, a_t)})p(\mathbf{y}_t^{a_t}|e_t, \mathbf{Y}^{t-1}), \end{aligned} \quad (20)$$

where $R(e_t, a_t)$ is the set of indices of clutter measurements under the joint 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 targets and missed detections, *i.e.* $e_t = 0$ and $a_t = 0$. Thus, the only unknown is the prediction for the existence variable, which can be obtained by introducing the previous existence variable and using the law of total probability, *i.e.*

$$p(e_t|\mathbf{Y}^{t-1}) = \sum_{e_{t-1}} p(e_t|e_{t-1})p(e_{t-1}|\mathbf{Y}^{t-1}). \quad (21)$$

Substituting from the modelling equations this becomes

$$p(\bar{E}_t|\mathbf{Y}^{t-1}) = (1 - P_b)(1 - P_{E_{t-1}}) + P_d P_{E_{t-1}} \quad (22)$$

$$p(E_t|\mathbf{Y}^{t-1}) = P_b(1 - P_{E_{t-1}}) + (1 - P_d)P_{E_{t-1}}. \quad (23)$$

In practice it is necessary to enumerate all the valid existence-association pairs, and evaluate the expression in (20) for each of these 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 target is undetected, and $a_t = j > 0$ if the j -th measurement originated from the target. Once the joint existence-association probabilities are computed the required marginal and conditional probabilities can be straightforwardly obtained as

$$p(E_t|\mathbf{Y}^t) = \sum_{j=0}^{M_t} p(E_t, a_t = j|\mathbf{Y}^t) \quad (24)$$

$$\beta_j = p(a_t = j|E_t, \mathbf{Y}^t) = \frac{p(E_t, a_t = j|\mathbf{Y}^t)}{p(E_t|\mathbf{Y}^t)}. \quad (25)$$

3.3 A PDAF implementation

To conclude this section we present below an algorithmic summary of the operations at time t of the single target PDA tracking algorithm.

Algorithm 1: Single Target Tracking with Existence

- **Input:** $p(\mathbf{x}_{t-1}|E_{t-1}, \mathbf{Y}^{t-1})$ and $P_{E_{t-1}} = p(E_{t-1}|\mathbf{Y}^{t-1})$.
- Compute the state prediction distribution $p(\mathbf{x}_t|E_t, \mathbf{Y}^{t-1})$, as in (13).
- Compute the measurement prediction distribution $p(\mathbf{y}_t^*|E_t, \mathbf{Y}^{t-1})$, as in (9).
- Gate the measurements using the measurement prediction distribution $p(\mathbf{y}_t^*|E_t, \mathbf{Y}^{t-1})$.
- For the remaining measurements, enumerate all the valid joint existence-association hypotheses $\{e_t, a_t\}$.
- For each of the valid joint hypotheses, evaluate the joint posterior probability $p(e_t, a_t|\mathbf{Y}^t)$, as in (20), and normalise.
- Compute the marginal existence posterior $p(E_t|\mathbf{Y}^t)$, as in (24), and the conditional association posteriors $\beta_j = p(a_t = j|E_t, \mathbf{Y}^t)$, $j = 0, 1 \dots M_t$, as in (25).
- Compute the updated filtering distribution $p(\mathbf{x}_t|E_t, \mathbf{Y}^t)$. The general form for the update is given in (10). For a mixture implementation the general form becomes (17).
- Perform mixture reduction on the filtering distribution to keep the number of components manageable.

- **Output:** $p(\mathbf{x}_t|E_t, \mathbf{Y}^t)$ and $P_{E_t} = p(E_t|\mathbf{Y}^t)$.

As we have stressed throughout this paper the framework described 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 in (18) 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.

4 Multi-Target Tracking and Existence

For multi-target tracking the distribution of interest is the joint posterior of all the target 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 targets that can be tracked. In a sense K is the number of “slots” available for tracked targets, *i.e.* when a target is killed its slot may be arbitrarily reassigned to a newly detected future target.

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

$$p(\mathbf{x}_t^{1:K}, E_t^{1:K}|\mathbf{Y}^t) = \prod_{k=1}^K p(\mathbf{x}_t^k|E_t^k, \mathbf{Y}^t)p(E_t^k|\mathbf{Y}^t), \quad (26)$$

and it will be our objective to maintain these marginal distributions. For each individual target the state filtering proceeds in exactly the same manner as discussed in Section 3.1. However, the computation of the existence and association probabilities cannot be done independently for each of the targets. Since the origins of the measurements are unknown the existence and association variables for the individual targets are coupled, and have to be treated jointly. Thus, we need to consider the joint posterior of the existence and association variables for all the targets, *i.e.* $p(e_t^{1:K}, a_t^{1:K}|\mathbf{Y}^t)$. Once this joint is known the required marginals and conditionals for the individual targets are easily obtained by summation.

Using computations similar to those leading to (20) the required joint is easily shown to be

$$\begin{aligned} p(e_t^{1:K}, a_t^{1:K}|\mathbf{Y}^t) &\propto p(M_t^C)p_C(\mathbf{y}_t^{R(e_t^{1:K}, a_t^{1:K})}) \\ &\times \prod_{k=1}^K \left[p(a_t^k|a_t^{1:k-1}, e_t^k)p(e_t^k|\mathbf{Y}^{t-1})p(\mathbf{y}_t^{a_t^k}|e_t^k, \mathbf{Y}^{t-1}) \right], \end{aligned} \quad (27)$$

where we have made use of the factorised form of the association prior.

Again it is necessary to enumerate all the valid joint existence-association hypotheses. Whereas this is tractable for a single target, the computational complexity for a direct enumeration of the joint hypotheses in the multi-target case grows super-exponentially in the number of targets. For K targets there are 2^K possibilities for the joint existence variable $e_t^{1:K}$. For each of these the number of association hypotheses is factorial in the number of measurements M_t and active targets $K_t = \sum_{k=1}^K e_t^k$. Thus, it is clear that more efficient computational strategies are required.

Measurement gating is a crucial first step to reduce the number of hypotheses. But for a large number of targets and measurements gating alone is not sufficient. We will briefly outline below three alternative strategies to compute the joint existence-association probabilities without explicitly enumerating all the valid hypotheses.

4.1 Independent Groups

After gating it is possible to group targets such that targets in a particular group share measurements only with other targets in the same group. Targets that are close in the state-space will tend to be grouped together, whereas a well-separated target will tend to form its own group. If there are N_G groups, with G_i the set of indices of the targets in the i -th group, the joint existence-association posterior can be factorised as

$$p(e_t^{1:K}, a_t^{1:K}|\mathbf{Y}^t) = \prod_{i=1}^{N_G} p(e_t^{G_i}, a_t^{G_i}|\mathbf{Y}^t), \quad (28)$$

where each of the factors are of the form in (27). Thus, groups can be treated independently, leading to substantial computational savings if the maximum number of targets in any group is sufficiently smaller than the total number of targets.

4.2 Efficient Hypothesis Management

Efficient Hypothesis Management (EHM) has been introduced in [11] as a method to efficiently compute the marginal association probabilities without explicitly enumerating the valid hypotheses, in the case where all the targets are assumed to exist. It can be reasonably straightforwardly extended to cope with the possibility of targets not existing.

EHM essentially ignores² the first two terms in (27), and then exploits the fact that the factorisation over the targets can be represented by a tree, with one level for each target. The number of children for a particular node in the tree is determined by the number of valid associations for the target in the next level, given the associations for the path from the root up to that node in the tree.

²It may be possible to devise (approximate) strategies to somehow factor the prior for the number of clutter measurements and their likelihood over the targets.

For a large number of targets the tree representation contains many repeated sub-structures that can be combined, leading to a compact graph representation. The marginal association probabilities can then be computed by an upward-downward algorithm, leading to a very efficient strategy that does not require the enumeration of the hypotheses.

4.3 An importance sampling implementation

We can use importance sampling to obtain a Monte Carlo approximation for the existence-association posterior in (27). If we constrain the importance distribution to take the form

$$q(e_t^{1:K}, a_t^{1:K} | \mathbf{Y}^t) = \prod_{k=1}^K q(e_t^k, a_t^k | a_t^{1:k-1}, \mathbf{Y}^t), \quad (29)$$

the samples can be generated sequentially. This is very efficient, since the support for each target in the above factorisation is at most as large as that of a single target treated independently. The importance weight associated with a sample generated in this way should be set to

$$w_t \propto \frac{p(e_t^{1:K}, a_t^{1:K} | \mathbf{Y}^t)}{q(e_t^{1:K}, a_t^{1:K} | \mathbf{Y}^t)}. \quad (30)$$

Typically one would set the individual target proposals to

$$q(e_t^k, a_t^k | a_t^{1:k-1}, \mathbf{Y}^t) \propto p(a_t^k | a_t^{1:k-1}, e_t^k) \times p(e_t^k | \mathbf{Y}^{t-1}) p(\mathbf{y}_t^{a_t^k} | e_t^k, \mathbf{Y}^{t-1}), \quad (31)$$

so that the importance weight reduces to

$$w_t \propto p(M_t^C) p_C(\mathbf{y}_t^{R(e_t^{1:K}, a_t^{1:K})}). \quad (32)$$

With a sufficiently large number of these samples accurate Monte Carlo estimates can be computed for the required existence and association probabilities.

Note that the strategies proposed above are not mutually exclusive, but can be combined. More specifically, one would generally first find the independent groups, and then perform EHM or importance sampling on groups for which the number of targets do not allow direct enumeration.

5 Application

In this section we demonstrate the performance of the proposed framework on a challenging multi-target tracking problem. Our objective is to track football players in video sequences. Due to player and camera movement the number of targets (football players) in view varies continuously.

The measurement process involves a detector that scans the images for plausible regions that may contain football players. The detector returns a set of bounding boxes, one for each plausible region, that are parameterised by their 2D

location in the image and their scale relative to a reference bounding box. A full description of the detector is beyond the scope of this paper. It basically performs a multi-scale gridding of the image, and computes a matching score for each grid location with a reference appearance model based on colour and shape. Grid locations with a score above a predefined threshold are deemed plausible regions, and are further refined by Mean Shift optimisation [12].

Figure 1 depicts ten keyframes from a typical football sequence, along with the output of the detector. The objective here is to track all the football players in white. It is clear that the number of players in view changes at regular intervals throughout the sequence. Note also that the detector is not perfect, *i.e.* false positives (clutter measurements) and missed detections are plentiful.

We define the state of each target as its 2D location and scale in the image, along with the velocities of these quantities. Thus, the relationship between the target state and the detection is linear. For the target dynamics we adopt a near constant velocity model [3] 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 targets can be represented by a Gaussian mixture, and that the update to compute the new mixture components in (17) can be performed by the Kalman filter.

We initialise the algorithm with no active targets. For each frame new target are initialised on all measurements that are not associated with any of the existing targets. 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 targets are only verified to be true targets 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 target 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 targets leave the scene measurements disappear, and the existence probability decreases. Once this probability falls below 0.1 we deem the target to have left the scene, and deactivate the corresponding tracker. For active targets we perform mixture reduction using the strategy described in [10], and we take target estimates to be the MAP value of the resulting Gaussian mixture.

A summary of the tracking results is presented in Figures 2 to 4. The complete videos are also available at the URL www.sigproc.eng.cam.ac.uk/~jv211/videos/exist. 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. Even though targets are wrongly initialised on clutter measurements they are never verified by the algorithm,

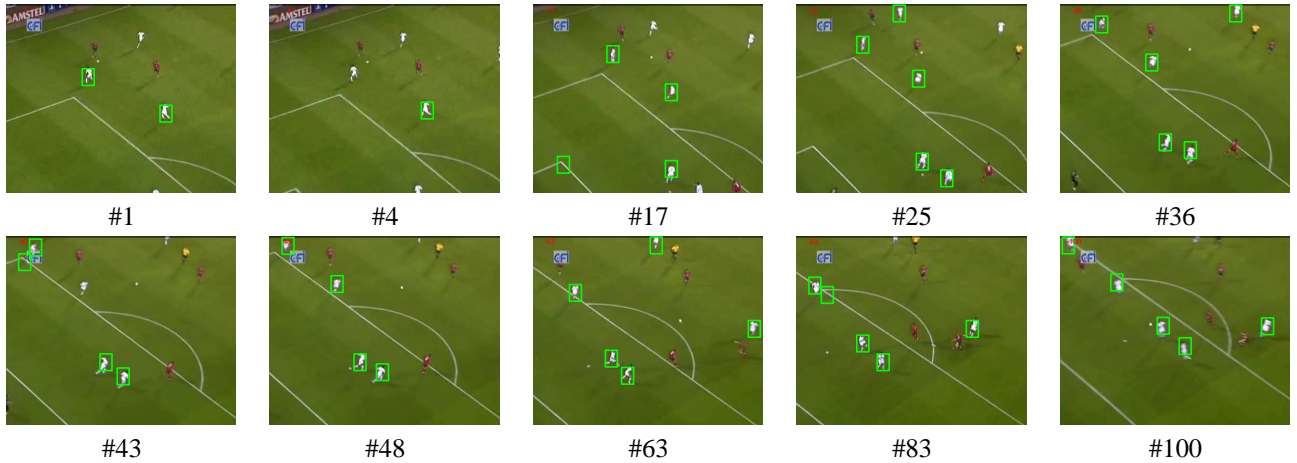


Figure 1: **Keyframes and detections.** The number of football players in view of the camera changes at regular intervals over the sequence. False positives (clutter) and missed detections abound.

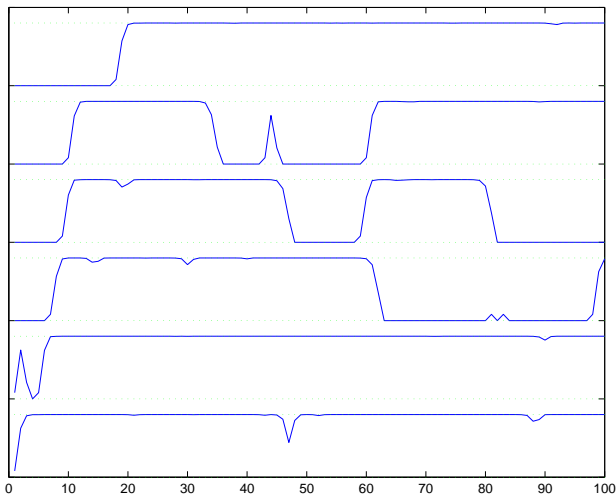


Figure 3: **Existence probabilities.** The existence probabilities for true targets soon ramp up to unity, whereas those for wrongly initialised targets fall to zero.

and the corresponding trackers are deactivated after a few time steps. The algorithm is able to smooth the location and scale estimates provided by the detector, and to bridge over missed detections without losing track.

For the results above we have computed the joint existence-association probabilities by first dividing the targets into independent groups, and then explicitly enumerating all the hypotheses for each group. In the majority of cases groups consisted of single targets. The other strategies (full joint enumeration and importance sampling) gave comparable results in terms of estimation accuracy, but differed in computational complexity. The comparative times, averaged over ten runs for each strategy, are depicted in Figure 5. From this figure the superiority of the indepen-

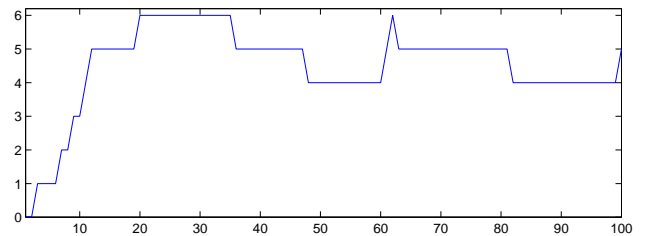


Figure 4: **Number of targets.** The estimate for the number of active targets accurately reflect the information in the video sequence.

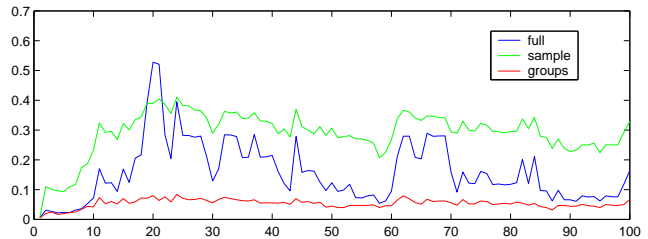


Figure 5: **Computation time in seconds.** Independent grouping is the superior strategy. Importance sampling will outperform full enumeration for a larger number of targets.

dent grouping strategy is evident. Importance sampling (50 samples were used) was outperformed by full enumeration, but this situation will be quickly reversed if the maximum number of targets increases any further.

6 Conclusions

We have presented a unifying framework for the joint treatment of target tracking and existence generalising a plethora of publications within the tracking literature. The

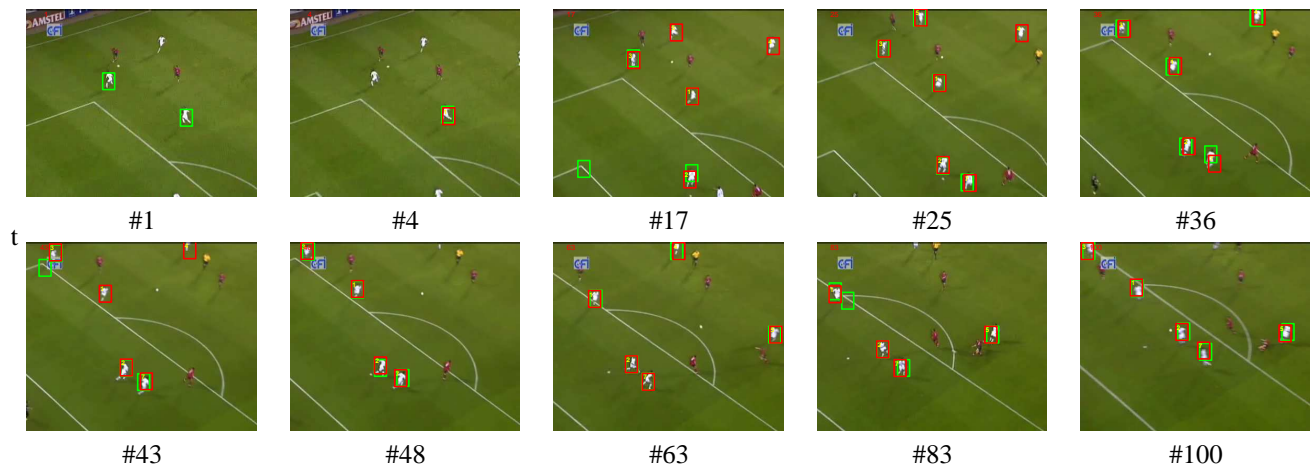


Figure 2: **Tracking results.** The algorithm successfully detects and tracks all football players for as long as they remain in view of the camera system.

framework is able to robustly deal with an unknown and variable number of targets by efficiently computing the posterior distribution of the target existence and association variables by being invariant to the inference/data association mechanism employed (thus allowing the “best tool for the job at hand” to be utilised). We have applied the algorithm to the difficult problem of detecting and tracking football players in video sequences, where we have adopted a mixture Kalman filter implementation of the general framework. Despite a constantly changing number of targets and a large number of missed detections and false positives the algorithm was able to successfully detect and track all the football players for as long as they are within the view of the camera system. Future work will focus on implementations for different applications (for example the notion of group existence, and then target within group existence), and extensions to cope with a very large number of targets and sensors.

7 Acknowledgements

The authors would like to thank the Royal Commission for the Exhibition of 1851 (www.royalcommission1851.org.uk). This research was supported by the UK MOD Corporate Research Programme through work funded by the Defence Technology Centre for Data and Information Fusion.

References

- [1] D. Mušicki, R. Evans, and S. Stanković. Integrated probabilistic data association. *IEEE Transactions on Automatic Control*, 39(6):1237–1241, June 1994.
- [2] D. Mušicki and R. Evans. Joint integrated probabilistic data association – JIPDA. In *Proceedings of*

the Fifth International Conference on Information Fusion, volume 2, pages 1120–1125, 2002.

- [3] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.
- [4] T. Kirubarajan and Y. Bar-Shalom. Probabilistic data association techniques for target tracking in clutter. *Proceedings of the IEEE*, 92(3):536–557, March 2004.
- [5] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8:173–184, 1983.
- [6] J. Vermaak, S. J. Godsill, and P. Pérez. Monte Carlo filtering for multi-target tracking and data association. 2005. Accepted for publication in the *IEEE Transactions on Aerospace and Electronic Systems*.
- [7] M. Rollason and D. Salmond. A particle filter for track-before-detect of a target with unknown amplitude. In *Target Tracking: Algorithms and Applications (Ref. No. 2001/174)*, *IEE*, volume 1, pages 14/1–14/4, 2001.
- [8] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proceedings of AeroSense: The 11th International Symposium on Aerospace / Defence Sensing, Simulation and Controls*, volume Multi Sensor Fusion, Tracking and Resource Management II, 1997.
- [9] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

- [10] D. J. Salmond. Mixture reduction algorithms for target tracking in clutter. In O. E. Drummond, editor, *Signal and Data Processing of Small Targets, SPIE 1305*, pages 434–445, 1990.
- [11] S. Maskell, M. Briers, and R. Wright. Fast mutual exclusion. In *Proceedings of SPIE*, 2004.
- [12] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages II: 142–149, 2000.