The Identity Management Problem — A Short Survey

Leonidas J. Guibas
Computer Science Department
Stanford University
Stanford, CA 94305, U.S.A.
Email: guibas@cs.stanford.edu

***Abstract*—The identity management problem is the problem of probabilistically keeping track of the association between target tracks and target identities, based on observations made by sensors. Updates of the belief state can happen because of new sensor observations reflecting on target identity, or because targets come near each other so that their identities become confused or mixed. Since the space of all possible associations grows factorially with the number of targets, it becomes important to find compact representations for distributions over associations and efficient implementations of the associated filter operations. In this note we introduce and describe the identity management problem and then place in context and briefly survey some the earlier work on this problem by us and others.**

Keywords: Tracking, identity management, data association, representation theory on groups.

I. Introduction

In this short note we introduce and discuss the *identity management problem* as it arises in the context of tracking multiple targets while fusing data from multiple distributed sensors. We summarize some of the recent work on the problem and give pointers to the literature for further reading. This note is based on joint work with several collaborators, including Carlos Guestrin, Jonathan Huang, Nelson Lee, Kunle Olukotun, Brad Schumitsch, Jaewon Shin, Sebastian Thrun, and Feng Zhao.

Suppose we are in a setting where a sensor network is tasked with tracking multiple, simultaneously moving targets, say vehicles. Tracking is of course a very well studied problem and many classical tools exist for representing and tracking positional uncertainty in the targets, such as Kalman filters [1], [2], particle filters [3]–[6], or mixtures of both (through 'Rao-Blackwellized particle filters,' see [7]). Here, however, we are primarily interested in certain *discrete aspects* of the problem, namely in the tracking of the identities of the moving targets so that we can distinguish the targets from each other and decide 'who is who.'

When the vehicles are well separated, the problem factorizes nicely and different sets of sensor nodes can focus on different vehicles, forming collaboration groups to best determine target positions. When vehicles pass near each other, however, confusion can arise as their signal signatures may mix; see Figure 1. After the vehicles separate again, their positions may be clearly distinguishable, but their identities can still be confused, as the sensors may no longer be able to tell which target is which. This identity uncertainty has to be carried forward in time with each vehicle, until additional sensing allows disambiguation. But how to accomplish this is subtle — when the identity of vehicle A becomes unambiguous due to an observation local to vehicle A, another vehicle B with which A had earlier 'mixed' may also become unambiguous. But currently B may be at some distant location, so how is the sensor network to resolve this ambiguity globally? We call this the *identity management problem* — the task of maintaining a belief state for the correct association between vehicle tracks and vehicle identities under local mixing events and sensor observations.

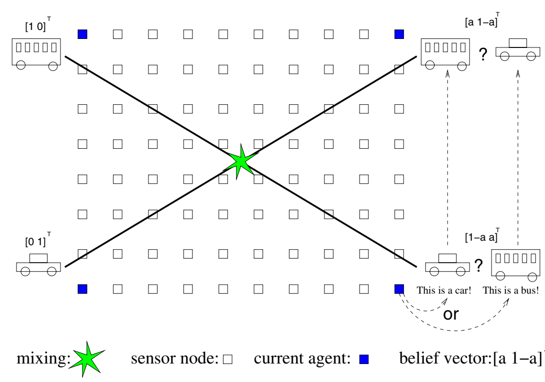

Figure 1. When vehicles pass near each other, their identities can get confused.

The identity management problem is closely related to, but not identical with, the classical data association problem of maintaining correspondences between tracks and observations. In the identity management problem, the rate at which observations happen that are informative about vehicle identities is not coupled to the rate of observations about vehicle positions and can be a lot lower. Both problems need to address the fundamental combinatorial challenge that there is a factorial or exponential number of associations to maintain between tracks and identities, or between tracks and observations respectively. The identity management problem is hard even in a centralized setting where the cost of communicating sensor data is not an issue.

A vast literature already exists on the the data association problem, beginning with the multiple hypothesis testing approach (MHT) of Reid [8]. MHT is a 'deferred logic' method in which past observations are exploited in forming new hypotheses when a new set of observations arises. Since the number of hypotheses can grow exponentially over time, various heuristics have been proposed to help cope with this complexity. For example, one can choose to maintain only the k best hypotheses for some parameter k [9], using Murty's algorithm [10]. But for this approximation to be effective, k still has to be exponential in the number of objects n. A slightly more recent filtering approach is the joint probabilistic data association filter (JPDA) [11], which is a suboptimal single-stage approximation of the optimal Bayesian filter. JPDA makes associations sequentially and is unable to correct erroneous associations made in the past [12]. Even though it is more efficient than MHT, the calculation of the JPDA association probabilities is still an NP-hard problem [13]. Polynomial approximation algorithms to the JPDA association probabilities have recently been studied using Markov chain Monte Carlo (MCMC) methods [14], [15].

As mentioned above, the key computational challenge in the identity management problem is that the number of possible associations between tracks and target identities can get very large. Even in the simplest 'closed world' setting, this is the difficult problem of performing inference on distributions over the symmetric group S_n describing all possible associations between n tracks and n object identities. The size of S_n is $n!$ and thus an explicit representation of such a distribution becomes unmanageable even for very modest n.

II. The Belief Matrix Approach

The identity management problem was first explicitly introduced in [16]. It differs from the classical data association problem in that our observation model is not concerned with the low-level tracking details but instead with high level information about object identities. In that paper we introduced the notion of the *belief matrix* approximation of the association probabilities, which collapses an exponential distribution over all possible associations to just its first-order marginals. In the simplest case of n tracks and n identities, the belief matrix B is an $n \times n$ doubly-stochastic matrix of non-negative entries b_{ij}, where b_{ij} is the probability that identity i is associated with track j. The double stochastic conditions that all rows and columns sum to 1 form the natural normalization conditions on these probabilities (i.e., that each track corresponds to some ID, and vice versa). Note that by simply knowing the n^2 entries of the belief matrix, we are able to probabilistically answer some basic identity questions, such as

- show the location (track) of an object with a given identity
- show the identity of an object at a given location (track)

As the system evolves forwards in time, the belief matrix needs to be updated. The association between identities and tracks can change either when there is new sensor evidence about the identity of a particular track (because some measured property of an object makes it more likely to be a particular object), or when two objects come near each other so that their locations and properties become hard to properly associate (as in the vehicle example presented earlier). We call the former identity *evidence events* and the latter identity *mixing events*. Mixing events are easy to capture in the belief matrix representation: the two columns of B that represent the tracks in question get blended with certain coefficients. Evidence, events, however, are harder. The reason is that we cannot just reapportion identity mass within the entries of a single column of B (corresponding to the observed track), without also worrying about preservation of the double stochasticity constraints among the rows.

Although in principle one could keep track of all past mixing and evidence events and compute a full Bayesian posterior, this is very expensive. In [16] we proposed a much simpler method that uses the Sinkhorn iteration or scaling [17] for renormalizing a belief matrix. This is an iterative procedure in which the rows and then the columns of B are repeatedly normalized independently, until convergence is obtained. Convergence is guaranteed under very mild conditions and it can be shown that the limit is a non-negative doubly stochastic matrix C that is closest to the modified B, in the sense that the Kullback-Leibler divergence or cross entropy [18] from C to B

$$I(B:C) = \sum_{j=1}^{n}\sum_{i=1}^{n} c_{ij} \log \frac{c_{ij}}{b_{ij}}$$

is minimized [19]. It can further be shown that this limit C shares many desirable properties with the true Bayesian posterior, in that it avoids introducing any biases or assumptions not present in the original data [16].

In a distributed sensor network setting, we would store columns of the belief matrix in leader nodes — typically sensor nodes close to each moving object. As objects move, the leadership may change hands among sensor nodes. Since motion continuous in nature, these hand-offs occur within a local geographic neighborhood. In this way, the state of the system can be maintained and updated in a distributed fashion, with state migration remaining a localized operation.

Mixings are easy to implement in this distributed belief matrix representation as the leaders holding info about the tracks being confused must be near in the network. Observations, however, require matrix re-normalization. On the one hand, column normalizations are easy local operations. Row normalizations, on the other hand, can be expensive, as all leader nodes holding a track that has a non-zero probability of being that row's particular identity need to participate. This is a type of sensor node collaboration group defined by the task needs and one in which the corresponding nodes may be non-local and indeed geographically widely separated. The communication pattern that arises is that of maintaining connectivity among a group of roaming agents — the leader nodes tracking those vehicles whose identities are potentially mixed. This requires multicast routing support at the network layer. In [20] we show a symmetric multicast routing protocol that connects a group of processes residing in sensor nodes and migrating from one node to another following physical events being tracked.

To summarize, in the belief matrix representation, mixing events are local computations, while evidence events may involve repeated and distant communication as different leader nodes perform repeated scalings for Sinkhorn to converge.

III. Band-limited Fourier Representations

It turns out that the belief matrix approach can be extended to higher order marginals such as, for example, the association probabilities between pairs of tracks and pairs of target identities, either in the unordered or ordered senses. However, the mathematics becomes quickly unmanageable, in part because the marginals of different orders are interconnected — for example, the second-order marginals imply the first-order marginals, and so on. Even among the first-order marginals, we have the double stochasticity conditions whose maintenance under sensor observations, as we saw above, is difficult.

There is, however, an established mathematical theory that is ideally suited to disentangling all this information: the *representation theory of the symmetric group* [21]–[23]. The same way that one can define Fourier transforms for functions defined over the reals, there is a corresponding way to define Fourier transforms for functions defined over groups, and in particular over the symmetric group S_n. A *representation* of a group G is a map ρ from G to a set of invertible $d_\rho \times d_\rho$ matrix operators which preserves algebraic structure in the sense that for all $\sigma_1, \sigma_2 \in G$, $\rho(\sigma_1\sigma_2) = \rho(\sigma_1)\cdot\rho(\sigma_2)$, in other words a homomorphism. The matrices which lie in the image of ρ are called the *representation matrices*, and we will refer to d_ρ as the *degree* of the representation.

We begin by showing three examples of representations on the symmetric group, S_n.

1) The simplest example of a representation is called the *trivial representation* $\rho_{(n)} : S_n \rightarrow \mathbb{R}^{1\times 1}$, which maps each element of the symmetric group to 1, the multiplicative identity on the real numbers. The trivial representation is actually defined for every group, and while it may seem unworthy of mention, it plays the role of the constant basis function in the Fourier theory.
2) The *first-order permutation representation* of S_n is the degree n representation, $\tau_{(n-1,1)}$, which maps a permutation σ to its corresponding permutation matrix given by $[\tau_{(n-1,1)}(\sigma)]_{ij} = \mathbb{1}\{\sigma(j) = i\}$. For example, the first-order permutation representation on S_3 is given by:

$$\tau_{(2,1)}(\epsilon) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \tau_{(2,1)}(1,2) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\tau_{(2,1)}(2,3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \tau_{(2,1)}(1,3) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$
$$\tau_{(2,1)}(1,2,3) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \tau_{(2,1)}(1,3,2) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

σ	$\rho_{(3)}$	$\rho_{(2,1)}$	$\rho_{(1,1,1)}$
ϵ	1	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	1
$(1,2)$	1	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	-1
$(2,3)$	1	$\begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}$	-1
$(1,3)$	1	$\begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}$	-1
$(1,2,3)$	1	$\begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}$	1
$(1,3,2)$	1	$\begin{bmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}$	1

Table I
THE IRREDUCIBLE REPRESENTATION MATRICES OF S_3.

3) The *alternating representation* of S_n, maps a permutation σ to the determinant of $\tau_{(n-1,1)}(\sigma)$, which is $+1$ if σ can be written as the composition of an even number of pairwise swaps, and -1 otherwise. We write the alternating representation as $\rho_{(1,\ldots,1)}$ with n 1's in the subscript. For example, on S_4, we have:
$$\rho_{(1,1,1,1)}((1,2,3)) = \rho_{(1,1,1,1)}((13)(12)) = +1.$$
The alternating representation can be interpreted as the 'highest frequency' basis function on the symmetric group, intuitively due to its high sensitivity to swaps. For example, if $\tau_{(1,\ldots,1)}(\sigma) = 1$, then $\tau_{(1,\ldots,1)}((12)\sigma) = -1$. In identity management, it may be reasonable to believe that the joint probability over all n identity labels should only change by a little if just two objects are mislabeled due to swapping — in this case, ignoring the basis function corresponding to the alternating representation should provide a extremely accurate approximation to the joint distribution.

In general, a representation corresponds to an overcomplete set of functions and therefore does not constitute a valid basis for any subspace of functions. For example, the set of nine functions corresponding to $\tau_{(2,1)}$ span only four dimensions, because there are six normalization constrains (three on the row sums and three on the column sums), of which five are independent. To find a valid complete basis for the space of functions on S_n, we will need to find a family of representations whose basis functions are independent, and span the entire $n!$-dimensional space of functions. These are called the *irreducible representations*.

As it happens, there are only three irreducible representations of S_3 [21], the trivial representation $\rho_{(3)}$, the degree 2 representation $\rho_{(2,1)}$, and the alternating representation $\rho_{(1,1,1)}$.

The complete set of irreducible representation matrices of S_3 are shown in Table I.

The Fourier transform (as defined below) at the trivial representation allows us to capture the zeroth-order marginal (normalization constant) of a distribution. By considering the trivial representation together with another irreducible of degree 2 (called $\rho_{(2,1)}$), it is possible to recover first-order marginal probabilities. The irreducible $\rho_{(2,1)}$ can be thought of as a basis for the set of first-order functions on S_3 whose zeroth order effects have been subtracted off (i.e. first-order functions that sum to zero). One can show that an analogous version of $\rho_{(2,1)}$ exists for any n that accounts for *pure* first-order effects. The reader is referred to [24] for further details and examples.

In general, the set of values at all irreducible representations provides a complete orthogonal basis for all functions over the symmetric group and defines the Fourier transform. More formally, let $f : G \to \mathbb{R}$ be any function on a group G and let ρ be any representation on G. The *Fourier Transform* of f at the representation ρ is defined to be the matrix of coefficients:

$$\hat{f}_\rho = \sum_\sigma f(\sigma)\rho(\sigma).$$

The collection of Fourier Transforms at all irreducible representations of G forms *the Fourier Transform of* f.

As in the familiar formulation, the Fourier Transform is invertible and the inversion formula is explicitly given by the Fourier Inversion Theorem.

$$f(\sigma) = \frac{1}{|G|} \sum_\lambda d_{\rho_\lambda} \mathrm{Tr}\left[\hat{f}^T_{\rho_\lambda} \cdot \rho_\lambda(\sigma)\right],$$

where λ indexes over the collection of irreducibles of G.

Note that the trace term in the inverse Fourier Transform is just the 'matrix dot product' between $\hat{f}_{\rho_\lambda}$ and $\rho_\lambda(\sigma)$, since $\mathrm{Tr}\left[A^T \cdot B\right] = \langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle$, where by vec we mean mapping a matrix to a vector on the same elements arranged in row-major order.

The key idea is to represent distributions over permutations compactly, by band-limiting their Fourier spectra, in the hope that discarding higher order terms does not lose too much useful information about the distribution. Of course our main goal is to maintain such distributions as mixing events happen (which increase the distribution entropy) as well observation events (which decrease the entropy). The approach would be pointless, however, if we were forced to invert from the Fourier representation of a distribution back to the primal representation in order to perform these inference operations. Naively, the Fourier Transform on S_n scales as $O((n!)^2)$, and even the fastest Fast Fourier Transforms for functions on S_n are no faster than $O(n^2 \cdot n!)$ [25].

These ideas were first explored by Willsky [26], where the probabilistic filtering/smoothing problem for group-valued random variables was formulated. He provided an efficient FFT-based approach of transforming between primal and Fourier domains so as to avoid costly convolutions, and provided efficient algorithms for dihedral and metacyclic groups. Later, Kueh et. al. [27] showed that probability distributions on the group of permutations are well approximated by a small subset of Fourier coefficients of the actual distribution, allowing for a principled tradeoff between accuracy and complexity. Most recently, Kondor et. al [28] allowed for a general set of Fourier coefficients, but assumed a restrictive form of the observation model in order to exploit an efficient FFT factorization.

In [24] we show a new and simple algorithm, called *Kronecker Conditioning*, which performs all probabilistic inference operations completely in the Fourier domain, which allows for a principled tradeoff between computational complexity and approximation accuracy. The approach is fully general, in the sense that it can address any transition model or likelihood function that can be represented in the Fourier domain, such as those used in previous works, and can represent the probability distribution using any desired number of Fourier coefficients. We have analyzed the errors which can be introduced by bandlimiting a probability distribution and showed how they propagate with respect to inference operations. Approximate conditioning based on bandlimited distributions can on occasion yield Fourier coefficients which do not correspond to any valid distribution, sometimes even returning negative probabilities — we have addressed this issue by presenting a method for projecting the result back into the polytope of coefficients which correspond to nonnegative and consistent marginal probabilities using an efficient quadratic program. The interested reader is referred to [24] for the details.

IV. The Identity Matrix Approach

An alternative representation that has also been considered is an information-theory based approach [18]. In this case we consider an *information matrix* Ω whose elements represent the amount of information we have that a certain track corresponds to a certain identity. Each entry is an additive term in a *negative log-likelihood* function. If, for example, we want to know the actual likelihood of a concrete permutation matrix, we extract the corresponding elements of our sparse information matrix, add them up, exponentiate (to get rid of the logarithm), and normalize. Though normalization may again be difficult, we can easily compute ratios of probabilities between different permutation matrices exactly, because the normalizing factor cancels. Thus, the sparse information matrix is an efficient summary of the identity information that otherwise would require exponential space. This differs from the MHT approach discussed above, which samples from the space of identity matrices (and hence scales exponentially). The information matrix approach instead maintains a single matrix that caches away the available identity information and requires memory at most equal to the product of the number of tracks and identities. The information matrix approach is especially attractive in a distributed sensor network setting as, if the columns of the information matrix are distributed to leader nodes tracking the respective targets, then the observation events become entirely local operations, avoiding the expensive renormalizations required in the belief matrix representation. On the other hand, mixing events become more complicated in the information matrix form. As in many classical data structures problems there are representation trade-off issues: some operations are less expensive in one representation and others in the the other. The best choice in any particular scenario will depend on the ratio between observation and mixing events. The details of the information matrix approach have been developed in [29]–[31].

Although the Fourier approach outlined earlier presents a principled way to extend the belief matrix approach to marginals of arbitrary order, we do not know of a comparable solution to the information matrix approach — though in principle it should exist.

In principle, the identity management problem and the data association problem are not independent. For example, information about object identities can influence data association decisions (e.g., a heavy moving vehicle cannot turn so fast) or, conversely, motion data can help identify objects (e.g, slow motion might indicate that a person is wounded or elderly). Much remains to be done, however, to develop a framework where all these dependencies can be properly accounted for, or to deal with dynamic settings allowing for both object insertion and deletion, as well for intermittently observable objects.

As a first step towards this integration between the discrete identity management problem and the more classical continuous tracking of positions, [31] proposes the *identity management Kalman filter* (IMKF), which maintains over the tracks a conventional positional mean μ and covariance Σ, but also a novel quadratic data association matrix Ω. The IMKF bypasses the problem arising from unknown data associations by maintaining internal "tracks." Each such track might correspond to multiple objects in the environment, at different points in time. The probabilistic correspondence between the internal tracks and the objects in the environment is encoded in the identity association matrix Ω. As described above, this matrix maintains a posterior in information form, which is an unnormalized logarithmic form of the data association probabilities. The advantage of this form is that (under mild approximations) it encodes the otherwise exponential distribution over the exponentially many ways to assign internal tracks to objects in the physical world, into a much simple quadratic matrix. In deriving the IMKF, we made a number of approximations to escape the otherwise inherent exponential complexity of the true posterior. The key approximation lies in the fact that we use a maximum likelihood data association track for updating the continuous KF parameters; however, the discrete data association probabilities are then updated in the full Bayesian posterior form. Another approximation occurs when updating the data association matrix. Here we approximated a set of marginals using Jensen's inequality, which is essential for avoiding an exponential expansion of the posterior. While these approximations are significant, they appear not to harm the filter's ability to generate useful full posteriors over the objects being tracked under data association uncertainty. As experimental results illustrate, our filter effortlessly scales to large values of n.

In simplified form, the IMKF works as follows. The IMKF maintains n continuous tracks, each characterized by a mean μ_i and a covariance matrix Σ_i. In updating these tracks, the IMKF uses the classical maximum-likelihood data association method, which consists of simply finding the most likely assignment of observed objects to internal tracks. It then updates μ_i Σ_i using the standard Kalman filter equations.

Because the internal tracks may not be easily associated with individual objects in the physical world, the IMKF allows for a permutation mapping for assigning tracks to physical objects. An example permutation matrix is shown below. It associates internal tracks (columns) to object identities (rows); the bold-face entry **1** associates track #2 with object #1 in this example.

$$A \quad = \quad \begin{pmatrix} 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Each such permutation matrix is a valid data association. IMKF maintains implicitly an entire probability distribution over all permutation matrices of this type. This probability is encapsulated in yet another matrix, the *data association information matrix* Ω. Ω defines the probability of each permutation A in the following way:

$$p(A) \quad \propto \quad \exp\ \mathrm{trace}\,(\Omega \cdot A)\,.$$

This simple equation informally states: the more Ω 'agrees' with A, the more likely is the data association A.

In [29], [30], it is shown that all essential updates of Ω are *local* and *sparse*. For example, if a sensor observes that a tracked object 3 likely corresponds to a physical object 2, Ω is updated as follows:

$$\begin{pmatrix} y_{11} & y_{21} & y_{31} & y_{41} \\ y_{12} & y_{22} & y_{32} & y_{42} \\ y_{13} & y_{23} & y_{33} & y_{43} \\ y_{14} & y_{24} & y_{34} & y_{44} \end{pmatrix} \longrightarrow$$
$$\begin{pmatrix} y_{11} & y_{21} & y_{31} & y_{41} \\ y_{12} & y_{22} & y_{32}{+}C & y_{42} \\ y_{13} & y_{23} & y_{33} & y_{43} \\ y_{14} & y_{24} & y_{34} & y_{44} \end{pmatrix}$$

Here C represent the evidence event information that associates column 3 and row 2 and is a function of the sensor's reading confidence and noise level.

Further, if two objects come so close together in the physical world that they cannot be disambiguated, the IMKF uses yet another local update. This results in a mixing of the corresponding rows in Ω, of the following form (shown here for columns #2 and #3). Let 'exp' and 'log' denote element-wise exponentiation and logarithm for a matrix, and define $W = \exp \Omega$. Then the mixing is done in the exponentiated matrix form

$$\begin{pmatrix} w_{11} & w_{21} & w_{31} & w_{41} \\ w_{12} & w_{22} & w_{32} & w_{42} \\ w_{13} & w_{23} & w_{33} & w_{43} \\ w_{14} & w_{24} & w_{34} & w_{44} \end{pmatrix} \longrightarrow$$

$$\begin{pmatrix} w_{11} & \frac{1}{2}(w_{21}+w_{31}) & \frac{1}{2}(w_{21}+w_{31}) & w_{41} \\ w_{12} & \frac{1}{2}(w_{22}+w_{32}) & \frac{1}{2}(w_{22}+w_{32}) & w_{42} \\ w_{13} & \frac{1}{2}(w_{23}+w_{33}) & \frac{1}{2}(w_{23}+w_{33}) & w_{43} \\ w_{14} & \frac{1}{2}(w_{24}+w_{34}) & \frac{1}{2}(w_{24}+w_{34}) & w_{44} \end{pmatrix}$$

and the new Ω is just $\log W$, after the update.

This information matrix Ω is at the core of the IMKF approach for handling uncertain data associations. Its suitability for a distributed sensor network arises from two properties: If mixings are limited, as we expect in practice, the matrix Ω is sparse (it only uses constant space per sensor node); and all updates of Ω are local.

Again, the key benefit of the information matrix approach is that it makes evidence events completely local to the area of the network where the observations occur. The leader nodes simply accumulate log-likelihoods for the different identities and do not need to normalize these into actual probabilities, as was required in the belief matrix case. This ability to locally accumulate and integrate into the information matrix representation new identity evidence is what provides efficiency and scalability. This is in contrast to the exponential update time of conventional data association filters and enables the design of systems that easily scale to tracking a number of objects that is comparable to the number of nodes in the network.

V. Conclusions

In this brief survey we introduced the identity management problems and its relation to classical tracking and data association problems.

We remark that many interesting problems remain, including several already mentioned, such as those of handling new objects entering the scene or old objects leaving, as well as the development of higher-order version of the information matrix approach.

The Fourier approach itself suffers from Heisenberg's uncertainty principle: whenever the identities of the targets become rather certain, more and more Fourier coefficients will be required to represent this information. Thus the truncated Fourier series is best suited for scenarios where there is high uncertainty on target identities. We expect there should be a way to factorize uncertainty so that highly certain individual or group associations (or highly certain disassociations) can be pulled out of a global Fourier representation and represented compactly.

Acknowledgments: This work is supported in part by the Army Research Office under grant W911NF-06-1-0275 and the National Science Foundation under grants NeTS-NOSS 0626151 and TF 0634803. In addition to the co-authors listed at the beginning, we are grateful to Robert Hough and Emre Oto for helpful and insightful discussions.

References

[1] P. Swerling, "A proposed stagewise differential correction procedure for satellite tracking and prediction," RAND Corporation, Tech. Rep. P-1292, January 1958.

[2] M. Grewal and A. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*. Canada: John Wiley and Sons, Inc., 2001.

[3] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machine," *Journal of Chemical Physics*, vol. 21, pp. 1087–1091, 1953.

[4] J. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, 1998.

[5] A. Doucet, J. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York: Springer Verlag, 2001.

[6] S. Thrun, "Particle filters in robotics," in *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, 2002.

[7] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, Stanford, 2000, pp. 176–183.

[8] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. on Automatic Control*, vol. 6, pp. 843–854, 1979.

[9] I. Cox and S. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," in *International Conf. on Pattern Recognition*, 1994, pp. 437–443.

[10] K. Murty, "An algorithm for ranking all the assignments in order of increasing cost," *Operations Research*, vol. 16, pp. 682–687, 1968.

[11] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. Academic Press, 1988.

[12] A. Poore, "Multidimensional assignment and multitarget tracking," in *Partitioning Data Sets*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 1995, vol. 19, pp. 169–196.

[13] J. Collins and J. Uhlmann, "Efficient gating in data association with multivariate distributed states," *IEEE Trans. Aerospace and Electronic Systems*, vol. 28, 1992.

[14] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," in *Proc. of the IEEE International Conference on Decision and Control, Paradise Island, Bahamas*, 2004.

[15] S. Oh and S. Sastry, "A polynomial-time approximation algorithm for joint probabilistic data association," in *Proc. of the American Control Conference, Portland, OR*, 2005.

[16] J. Shin, L. Guibas, and F. Zhao, "A distributed algorithm for managing multi-target identities in wireless ad-hoc sensor networks," in *Proc. 2nd International Workshop on Information Processing in Sensor Networks (IPSN03)*. Palo Alto, CA: Springer, April 2003, pp. 223–238.

[17] U. Rothblum and H. Schneider, "Scaling of matrices which have prescpecified row sums and column sums via optimization," *Linear Algebra Appl.*, pp. 737–764, 1989.

[18] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 1991.

[19] H. Balakrishnan, I. Hwang, and C. J. Tomlin, "Polynomial approximation algorithms for belief matrix maintenance in identity management," in *Proceedings of the 43rd IEEE Conference on Decision and Control, Bahamas*, 2004.

[20] Q. Fang, J. Liu, L. Guibas, and F. Zhao, "Roamhba: Maintaining group connectivity in sensor networks," in *Proc. 3rd International Symposium on Information Processing in Sensor Networks (IPSN)*, 2004, pp. 151–160.

[21] P. Diaconis, *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.

[22] A. Terras, *Fourier Analysis on Finite Groups and Applications*. London Mathematical Society, 1999.

[23] B. E. Sagan, *The Symmetric Group*. Springer, April 2001.

[24] J. Huang, C. Guestrin, and L. Guibas, "Efficient inference for distributions on permutations," in *In Advances in Neural Information Processing Systems (NIPS)*, December 2007.

[25] R. Kondor, "$\mathbb{S}_n$ob: a C++ library for fast Fourier transforms on the symmetric group," 2006, Available at `http://www.cs.columbia.edu/~risi/Snob/`.

[26] A. Willsky, "On the algebraic structure of certain partially observable finite-state markov processes," *Information and Control*, vol. 38, pp. 179–212, 1978.

[27] K. Kueh, T. Olson, D. Rockmore, and K. Tan, "Nonlinear approximation theory on finite groups," Department of Mathematics, Dartmouth College, Tech. Rep. PMA-TR99-191, 1999.
[28] R. Kondor, A. Howard, and T. Jebara, "Multi-object tracking with representations of the symmetric group," in *AISTATS*, 2007.
[29] J. Shin, N. Lee, S. Thrun, and L. Guibas, "Lazy inference ob object identities in wireless sensor networks," in *Fourth International Conference on Information Processing in Sensor Networks*, Los Angeles, CA, 2005.
[30] B. Schumitsch, S. Thrun, G. Bradski, and K. Olukotun, "The information-form data association filter," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2005.
[31] B. Schumitsch, S. Thrun, L. Guibas, and K. Olukotun, "The identity management Kalman filter (imkf)," in *Proceedings of Robotics: Science and Systems*, Philadelphia, PA, USA, August 2006.