18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

The GFMT HPMHT Puzzle

Peter Willett
ECE Department, University of Connecticut
Storrs CT 06269
Email: willett@engr.uconn.edu

Tod Luginbuhl
Naval Undersea Warfare Center
Newport, RI 08241
Email: t.e.luginbuhl@ieee.org

Marcus Baum
ECE Department, University of Connecticut
Storrs CT 06269
Email: baum@engineer.uconn.edu

***Abstract*—Most target tracking algorithms work with data at the *plot* or *detection* level; that is, after an initial signal processing step of thresholding and centroiding that delivers point "hits" for data association and filtering. The GFMT (*general frequency modulation tracker*) and HPMHT (*histogram probabilistic multi-hypothesis tracker*), on the other hand, work directly with pixellated observation data, as on a focal plane array (FPA), meaning that the signal processing step is integral to the tracker. The HPMHT filled most of a statistical pit in the GFMT, but a perceptible hole remains. We discuss this here: If the problem can be modeled – and it can – then a solution that derives from this model ought to be available.**

I. Introduction

As opposed to large-target tracking (from the computer vision community, see for example [12]) in which target shape is an important clue, in small-target tracking algorithms the target-originated observation is assumed to be a post-signal processing "point". However, consider the data shown in figure 1: observations from the target (which relative to a pixel is actually a point) are both *spread* (or blurred, or smeared) according to a point-spread function (PSF), here modeled as a Gaussian-shaped filter with a standard deviation (σ) of two pixels, but are also *quantized* (or binned or pixellated) such that all optical energy falling within a pixel is simply totaled and assigned to the bin's center. A direct feed of focal-plane array (FPA) data, even thresholded data from the FPA's as represented in figure 1, to MHT, JPDAF and even MLPDA [1] would work suboptimally, due both to the loss of information from coarse quantization and from the collapse of the usual data association assumption that at most one measurement per scan can arise from each target.

On the other hand, there are algorithms that work directly with FPA data as in figure 1. The idea of estimating (holistically) based on observations that have been histogrammed is discussed in [15]. This was extended and became a dynamic target tracker (the GFMT) making use of PMHT ideas in [13], [14]. This was rebranded the (HPMHT) in [19], [20] with the important addition of a "data dependent prior". The data dependent prior is perhaps an artifact but a necessary one: a weakness of the GFMT is that each histogram (or spectral) "count" is treated as an observation, with the result that the effect of the process (the target motion model) becomes asymptotically too small. However, the data-dependent prior is itself an ad-hoc fix, and while its proportionality behavior as a function of the number of counts is certainly necessary, its exact form is arbitrarily chosen.

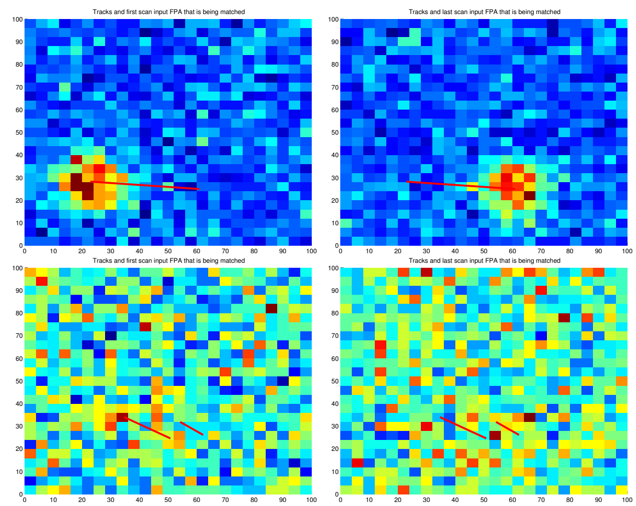

Fig. 1. Example of first (left plots) and last (right plots) scans of simulated focal plane array (FPA) data for straight line targets indicated by red lines. The upper row is moderate SNR and the lower is low-observable. This figure is intended that the reader understand the input data for the algorithms.

II. Models

A. Target Model

We assume T FPA scans (or frames) of data, and M targets that exist for all scans. The scans themselves are of dimension $I_1 \times I_2$ in terms of pixels and $a_1 \times a_2$ in terms of coordinate units. The m^{th} target's motion is assumed linear and perturbed by independent Gaussian *process noises* [2]:

$$x_m(t+1) \;=\; \mathbf{F}x_m(t) + w_m(t) \tag{1}$$

in which

$$\mathcal{E}\{w_m(k)w_m(k)^T\} \equiv \mathbf{Q} \tag{2}$$

Although it is not necessary we may as well assume the target's motion to be kinematic [1], [5], meaning that $x_m(t)$ comprises two dimensions each of position and velocity in the FPA coordinates; we will assume that $\mathbf{H}x_m(t)$ selects the two components of position. There is nothing in the GFMT, HPMHT or this discussion that requires two dimensions, invariant parameters or kinematic models; but to enlarge these assumptions would be dilatory.

B. Observation Model

The idea is that the FPA pixels measure optical energy; but that underlying this optical energy are the photons that gave rise to it. Let us assume that there are K photons – according to physics K ought to be a very large number indeed, and that is the point of this paper. Another necessary assumption is that the PSF be Gaussian in shape – the parameter $\mathbf{R}$ is a property of the optics, and it seems fair to assume that it be known in this development[1]. Let us further define probabilities $\{\pi_m\}_{m=0}^{M}$; in fact, π_m will be the probability that an arbitrarily-selected photon was in fact produced by target m, with $m = 0$ referring to clutter.

What is observable at the FPA is the aggregate count for each pixel. That is, the FPA observes

$$Z(t) = \{Z_i(t)\}_{i=1}^{I} \tag{3}$$

in which we are using the economical notations $i \equiv (i_1, i_2)$ & $I \equiv (I_1, I_2)$, and where

$$Z_i(t) = \sum_{k=1}^{K} \mathcal{I}\left(\mathcal{Q}\left[y_k(t)\right] = i\right) \tag{4}$$

in which the notation within the sum on the right-hand side (RHS) of (4) ($\mathcal{I}$ and $\mathcal{Q}$ are respectively indicator and quantization functions) means that a count is given to pixel $i = (i_1, i_2)$ at frame t when the k^{th} photon's location $y_k(t)$ lies within that pixel.

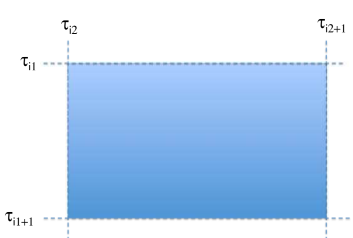

Fig. 2. Sketch of a pixel indicating thresholds. The subscripts "1" and "2" refer to the dimension.

In (4) we have used $y_k(t)$, the actual location of the k^{th} photon as it impinges on the t^{th} FPA frame. We have

$$y_k(t) \sim \begin{cases} \mathcal{N}(y; \mathbf{H}x_m(t), \mathbf{R}) & u_k(t) \in \{1, \ldots, M\} \\ \frac{1}{a_1 a_2} & u_k(t) = 0 \end{cases} \tag{5}$$

where the latter is uniform and $\mathcal{N}(y; \mu, R)$ denotes the Gaussian pdf in variable y with mean vector μ and covariance matrix $\mathbf{R}$, and in which $\{u_k(t)\}_{k=1}^{K}$ are drawn such that for $m = 0, 1, \ldots, M$

$$Pr(u_k(t) = m) \ = \ \pi_m \tag{6}$$

and are independent and identically distributed (iid). Clearly both $y_k(t)$ and $u_k(t)$ represent hidden variables of the sort familiar to those who work with the expectation maximization (EM) algorithm [10]. That is, we have

- $u_k(t) \in \{0, M\}$ denotes the provenance of the k^{th} photon at time t: it comes from one of the M targets or else clutter ($u_k(t) = 0$).
- $y_k(t) \in \Re^2$ denotes the true location in the FPA of the p^{th} photon at time t. Given $\mathcal{Q}\left[y_k(t)\right]$ it is restricted to $[\tau_{i1}, \tau_{i1+1}] \times [\tau_{i2}, \tau_{i2+1}]$, with reference to figure 2.

Since we will need them, we write their posterior probabilities in (8) and (9); the dependence on the EM iteration is suppressed. In (7), (8) and (9), $x_m(t)$ denotes the last (that is, $(n-1)^{st}$ EM iteration) estimate of the location of target m at scan t; and the limit of integration $y \in i$ implies integration over the i^{th} pixel.

III. The GFMT

The GFMT [13], [14] uses the models of the previous section, and applies EM [10] in the manner of the PMHT [4], [18], [23]. At fundament, the EM algorithm seeks to maximize

$$Q(X) = \int \log[p(Z, X, U, Y)] p(U, Y|Z) \tag{10}$$

where this "Q" should not be confused with process noise covariance $\mathbf{Q}$, and in which U, X, Y and Z denote the aggregated corresponding lower-case variables. We have

$$\begin{aligned} Q(X) \ &= \ \log[p(X)] + \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{m=0}^{M} P_t(m|\mathcal{Q}\left[y_k(t)\right]) \quad (11) \\ &\times \int_Y p_t(y_k(t)|\mathcal{Q}\left[y_k(t)\right], m) \log[\pi_m p(y_k(t)|x_m(t))] \end{aligned}$$

where P_t and p_t are from (8) and (9) and use the previous EM iteration's X, and in which it is important to recall that photons are being "counted" individually – we know $\mathcal{Q}\left[y_k(t)\right] \ \forall k \in \{1, K\}$. As in [23] we take the gradient with respect to X:

$$\begin{aligned} \nabla Q(X) \ &= \quad (12) \\ &\nabla \log[p(X)] + \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{m=1}^{M} P_t(m|\mathcal{Q}\left[y_k(t)\right]) \\ &\times \int_Y p_t(y_k(t)|\mathcal{Q}\left[y_k(t)\right], m) \mathbf{H}^T \mathbf{R}^{-1} [y_k(t) - \mathbf{H}x_m(t)] \end{aligned}$$

Then with

$$\begin{aligned} \bar{y}_{i,m}(t) \ &\equiv \ \mathcal{E}\left\{y_k(t)|\mathcal{Q}\left[y_k(t)\right] = i, u_k(t) = m\right\} \\ &= \ \int_{\mathcal{Q}[y]=i} y p_t(y|i, m) dy \quad (13) \end{aligned}$$

we get

$$\begin{aligned} \nabla Q(X) \ &= \ \nabla \log[p(X)] + \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{m=1}^{M} P_t(m|\mathcal{Q}\left[y_k(t)\right]) \\ &\times \mathbf{H}^T \mathbf{R}^{-1} [\bar{y}_{\mathcal{Q}[y_k(t)],m}(t) - \mathbf{H}x_m(t)] \quad (14) \end{aligned}$$

[1] For tractable operation it is probably necessary that $\mathbf{R}$ be diagonal.

$$P_t(i) \equiv Pr(\mathcal{Q}[y_k(t)]{=}i|x_m(t)) \quad = \quad \left(\pi_0 \int_{y\in i} \frac{1}{a_1 a_2} dy\right) + \sum_{l=1}^{M} \left(\pi_l \int_{y\in i} \mathcal{N}(y;\mathbf{H}x_l(t),\mathbf{R}) dy\right) \tag{7}$$

$$P_t(m|i) \equiv Pr(u_k(t){=}m|\mathcal{Q}[y_k(t)]{=}i,x_m(t)) \quad = \quad \begin{cases} \frac{\pi_m}{P_t(i)} \int_{y\in i} \mathcal{N}(y;\mathbf{H}x_m(t),\mathbf{R}) dy & m\in\{1,M\} \\ \frac{\pi_0}{P_y(i)} \int_{y\in i} \frac{1}{a_1 a_2} dy & m=0 \end{cases} \tag{8}$$

$$p_t(y|i,m) \equiv p(y_k(t)|u_k(t){=}m,\mathcal{Q}[y_k(t)]{=}i,x_m(t)) \quad = \quad \mathcal{I}(\mathcal{Q}[y]{=}i)\times \begin{cases} \frac{\mathcal{N}(y;\mathbf{H}x_m(t),\mathbf{R})}{\int_{y\in i} \mathcal{N}(y;\mathbf{H}x_m(t),\mathbf{R}) dy} & u_k(t){=}m\in\{1,M\} \\ \frac{\frac{1}{a_1 a_2}}{\int_{y\in i} \frac{1}{a_1 a_2} dy} & u_k(t){=}0 \end{cases} \tag{9}$$

The summation over k eventually counts through the $Z_1(t)$ photons that are in pixel $i = 1$, the $Z_2(t)$ photons that are in pixel 2, etc. Hence we can write (14) as

$$\nabla Q(X) \;=\; \nabla \log[p(X)] + \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{m=1}^{M} Z_i(t)P_t(m|i) \times \mathbf{H}^T\mathbf{R}^{-1}[\bar{y}_{i,m}(t) - \mathbf{H}x_m(t)] \tag{15}$$

or (it almost looks like a standard PMHT [4], [18], [23] now)

$$\nabla Q(X) = \nabla \log[p(X)] + K\sum_{t=1}^{T}\sum_{m=1}^{M} \mathbf{H}^T\tilde{\mathbf{R}}_m^{-1}(t)[\tilde{y}_m(t) - \mathbf{H}x_m(t)] \tag{16}$$

where

$$\tilde{\mathbf{R}}_m(t) \;\equiv\; \frac{\mathbf{R}}{\sum_{i=1}^{I} Z_i(t)P_t(m|i)} \tag{17}$$

$$\tilde{y}_m(t) \;\equiv\; \frac{\sum_{i=1}^{I} Z_i(t)P_t(m|i)\bar{y}_{i,m}(t)}{\sum_{i=1}^{I} Z_i(t)P_t(m|i)} \tag{18}$$

Note that $i \in \{1, I\}$ is shorthand for "all pixels," $\bar{y}_m(t)$ is from (13) and $P_t(m|i)$ is from (8). As with the usual PMHT explanation, $\nabla Q(X) = \nabla\breve{Q}(X)$ in which the latter corresponds to a standard linear/Gaussian model having "synthetic" measurements $\{\tilde{y}_m(t)\}$ with corresponding covariances $\{\tilde{\mathbf{R}}_m(t)\}$, or

$$\breve{Q}(X) = C - \frac{1}{2}\sum_{m=1}^{M}\Big[(x_m(1) - \bar{x}_m(1))^T\mathbf{P}_0^{-1}(x_m(1) - \bar{x}_m(1)) + \sum_{t=2}^{T}(x_m(t) - \mathbf{F}x_m(t-1))^T\mathbf{Q}^{-1}(x_m(t) - \mathbf{F}x_m(t-1)) + \sum_{t=1}^{T}(\tilde{y}_m(t) - \mathbf{H}x_m(t))^T\tilde{\mathbf{R}}_m^{-1}(\tilde{y}_m(t) - \mathbf{H}x_m(t))\Big] \tag{19}$$

in which $\mathbf{P}_0$ is the covariance of $x_m(t)$ at time $t = 1$. Accordingly, the EM iteration is via the Kalman smoother using these synthetic quantities.

Taking the gradient with respect to the π's and noting the constraint that they add to unity we have for $m = 0, 1, \ldots, M$

$$\pi_m = \kappa\sum_{t=1}^{T}\sum_{i=1}^{I} P_t(m|i) \tag{20}$$

where κ is chosen to make the sum unity. Note that while the π's in the usual PMHT implementation can be estimated, there is usually little reason for it and there may be numerical issues. However in the GFMT and HPMHT they are important and in general must be estimated: they represent the strength of the target.

Note also that in (20) the π's are fixed with time, reflecting (6-9). The modification is obvious. We also claim that the extension to observations of dimension not equal to two (say, one-dimensional "frequency bins" or three-dimensional voxels) is obvious. Some discussion of the multi-target situation, not so obvious, is in Appendix A.

IV. The HPMHT

Equation (17) suggests a concern: if the number of photons $K \to \infty$ then the synthetic measurement covariance $\tilde{\mathbf{R}}_m(t) \to 0$, since the denominator is proportional to K. To see this, suppose we were to normalize the FPA-level data such that (4) becomes

$$\tilde{Z}_i(t) = \frac{1}{K}\sum_{k=1}^{K}\mathcal{I}\left(\mathcal{Q}\left[y_k(t)\right] = i\right) \tag{21}$$

which sums to unity. Then (17) could be expressed as

$$\tilde{\mathbf{R}}_m(t) = \frac{\mathbf{R}}{K\sum_{i=1}^{I}\tilde{Z}_i(t)P_t(m|i)} \tag{22}$$

which makes the issue obvious.

We will discuss what this physically means in the next section; but operationally it means that the filtering becomes trivial, the measurements $\tilde{y}_m(t)$ tells us exactly where the target is at time t, and velocity can be inferred from their differences. Although the GFMT made its first appearance in [13], their simulations used finite values of K. The concern actually was noted in [14], but that followed its being pointed out in [19].

The histogram probabilistic multi-hypothesis tracker ([19], see also, among others, [6], [20]) adopts a different philosophy. Specifically, it is noted that the implied model (with hidden variables averaged out) for the GFMT is

$$p(X|Z) \;\propto\; p(X)\prod_{t=1}^{T}\prod_{k=1}^{K} p(y_k(t)|X) \tag{23}$$

then the effect of the "prior" $p(X)$ in the GFMT becomes *swamped* as in the large-K case by the observations. However, if the prior is made "data-dependent" such that

$$\mathbf{Q} \to \alpha \mathbf{Q}/K \tag{24}$$

in (19) then the effect is to replace $p(X)$ by $p(X)^K$: the swamping of the prior is avoided by increasing its effect proportionate to the number of photons per scan. The HPMHT uses $\alpha = 1$, and the resulting algorithm is essentially[2] the GFMT with $K = 1$ in (17).

V. The Puzzle

The GFMT's disenfranchisement of its prior can be understood in two ways.

- If $K \to \infty$, then for all pixels the number of photons diverges. The law of large numbers then renders the relative photon count (the photon count divided by K) deterministically equal to the expected count. From such a *noise-free* FPA the exact target positions is easily inferable with no error.
- The PMHT's data association model [18] allows for multiple measurements per target. Each photon is, in fact, a measurement, indistinguishable as far as model is concerned from a radar or sonar "hit" in the original PMHT, except with slightly reduced precision due to pixellation (i.e., quantization). If there are $K \to \infty$ photons then this is a PMHT with an infinite number of hits to work with.

From either perspective the amount of precision (in the sense of information or inverse variance) in the measurements is proportional to K; and that in the prior is fixed.

The HPMHT's *data-dependent* prior is a smart solution to the GFMT's *evanescent* one. With reference to (19) and noting that $\tilde{\mathbf{R}}_m(t)^{-1}$ is proportional to K it would seem that there really is no other choice: as $K \to \infty$ either the prior is evaporating, the data is ignored, or $\mathbf{Q} \propto 1/K$ and both prior and data are able to be involved in the filtering. The concern, however, is that one could replace $\mathbf{Q}$ by $\mathbf{Q}/K$, by $\mathbf{Q}/\pi$, even by $2000\mathbf{Q}/K$, and the same would be true. It would be much more satisfying if there were some physical modeling that could select the constant of proportionality.

The poser is that while the HPMHT is more appealing in its algorithmic form, the GFMT is derived from a model that is **correct according to the physics**[3]: photons from each point target are perturbed by the PSF from the optics; the photons are indeed captured and their powers aggregated in optical pixels; and there are so many photons that setting K to ∞ seems venial.

In the following section we offer some insight.

VI. Model-Driven Algorithms

A. Estimate K

It is above stated that the physics specifies that $K \to \infty$. In "bright" situations this is probably true, but in low-observable cases it may be a poor assumption: there may in fact be very few photons available in any pixel; or else the model is not well-matched to the FPA data. In such cases the match of observed level $Z_i(t)$ to its predicted level $KP_i(t)$ may be poor even after EM has done its optimization work. Actually, there are imaging devices that can work with very few photons such as the one presented in [16].

Hence an idea is to estimate K. It would be appealing to add K to the maximization of (11); but it makes makes little sense. According to the model K is known: from (3) we have the number of photons in each pixel, and the sum of these counts must be K. So even a poor match between $\{Z_i(t)\}$ and $\{KP_i(t)\}$ just a matter of repeated peculiar dice rolls.

So if we are operating in an environment where each photon or quantum can be counted, there is no reason to estimate K. On the other hand, many applications give $\{Z_i(t)\}$ in some other units – photons, volts, etc. And even though there would seem to be a mismatch between model (3) and observations $\{Z_i(t)\}$ that are not countable, the algorithm (Kalman smoother using (18) and (17)) remains quite applicable.

Let us define the normalized[4] pixel values as

$$\bar{Z}_i(t) \;\equiv\; \frac{Z_i(t)}{\sum_{i=1}^{I} Z_i(t)} \tag{25}$$

the match of which to $\{P_i(t)\}$ improves with K. We propose in this subsection iteratively to estimate K based on $\{\bar{Z}_i(T)\}$ and $\{P_i(t)\}$; then to reestimate $\{P_i(t)\}$ using K from the Kalman smoother that uses (18) and (17).

So the number of photons per pixel ($K\bar{Z}_i$ in pixel i, where it is assumed in the model that we have $K\bar{Z}_i$ is an integer) is multinomial, we can write

$$p(\{\bar{Z}_i(t)\}_{i=1}^{I}) = \frac{K!}{\prod_{i=1}^{I}(K\bar{Z}_i)!}\prod_{i=1}^{I}\left(P_i(t)\right)^{K\bar{Z}_i} \tag{26}$$

or, with a logarithm and allowing for non-integer $K\bar{Z}_i$ we have

$$\begin{aligned}\log\left[p(\{\bar{Z}_i(t)\}_{i=1}^{I})\right] &= \log\left[\Gamma(K+1)\right] \\ &+ \sum_{i=1}^{I}\left(K\bar{Z}_i\log\left[P_i(t)\right] + \log\left[\Gamma(K\bar{Z}_i+1)\right]\right)\end{aligned} \tag{27}$$

The answer is not explicit but it is a scalar maximization, so it is not especially onerous.

[2] It must be mentioned that the HPMHT makes another significant leap, albeit one orthogonal to the point of this paper. In the HPMHT the levels in the *unobserved* pixels beyond the edge of the FPA are afforded "hidden-variable" status in EM terms. The increase in algorithmic complexity is marginal, and the importance of this clever mechanism on bias-removal for targets whose PSF extends beyond the FPA should not be underestimated.

[3] Indeed, one could argue about photons versus particles, or whether it is justifiable to attribute each photon to a target according to (5), but the authors cannot think of a workable model that better describes the key ingredients to an FPA observations process than the one that the GFMT uses.

[4] We are normalizing to unity sum as was done in [22]. Any normalization will do, but one that is stochastic must have sum that depends on some stochastic variable that is exogenous to the observations model we are using.

In [22] it was suggested to model the $\bar{Z}_i(t)$'s as independent and Gaussian: $\mathcal{N}(z; P_i(t), P_i(t)(1-P_i(t)/K)$. We have

$$\hat{K} = \sum_{t=1}^{T} \frac{I}{\left(\sum_{i=1}^{I} \frac{\left(\bar{Z}_i(t)-P_i(t)\right)^2}{P_i(t)}\right)} \tag{28}$$

We note that in fact the $K\bar{Z}_i(t)$'s are jointly multinomial, and hence neither independent nor Gaussian, but the approximation seems often to work acceptably and is quick. With a finite K we can use the GFMT directly.

In either case (implicit maximization of (27) and explicit via (28)) the maximization can be done assuming that K is fixed for all t or may vary; the formulae are obvious.

B. Introduce Measurement Noise

Suppose (5) is replaced by

$$\begin{aligned} \gamma_m(t) &\sim \mathcal{N}(\gamma; \mathbf{H}x_m(t), \check{\mathbf{R}}) \qquad m = 1, 2, \ldots, M \\ y_k(t) &\sim \left\{ \begin{array}{ll} \mathcal{N}(y; \gamma_m(t), \mathbf{R}) & u_k(t) \in \{1, \ldots, M\} \\ \frac{1}{a_1 a_2} & u_k(t) = 0 \end{array} \right. \end{aligned} \tag{29}$$

We would then have (11) replaced by

$$\begin{aligned} Q(X, \Gamma) &= \log[p(X)] + \sum_{t=1}^{T}\sum_{m=1}^{M} \log[p(\gamma_m(t)|x_m(t))] \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{m=1}^{M} P_t(m|\mathcal{Q}\left[y_k(t)\right]) \\ &\times \int_Y p_t(y_k(t)|\mathcal{Q}\left[y_k(t)\right], m) \log[\pi_m p(y_k(t)|\gamma_m(t))] \end{aligned} \tag{30}$$

The similarity of (30) to (11) is striking. It is obvious that the prior on X and Γ (the first two terms on the RHS) becomes irrelevant as $K \to \infty$. We are eventually left with

$$\begin{aligned} &\check{Q}(X) = \\ &C - \frac{1}{2}\sum_{m=1}^{M} \left[(x_m(1) - \bar{x}_m(1))^T \mathbf{P}_0^{-1} (x_m(1) - \bar{x}_m(1)) + \right. \\ &\sum_{t=2}^{T} (x_m(t) - \mathbf{F}x_m(t-1))^T \mathbf{Q}^{-1} (x_m(t) - \mathbf{F}x_m(t-1)) \\ &\left. + \sum_{t=1}^{T} (\tilde{\gamma}_m(t) - \mathbf{H}x_m(t))^T \check{\mathbf{R}}^{-1} (\tilde{\gamma}_m(t) - \mathbf{H}x_m(t)) \right] \end{aligned} \tag{31}$$

in which

$$\tilde{\gamma}_m(t) \equiv \frac{\sum_{i=1}^{I} Z_i(t) \check{P}_t(m|i) \bar{y}_{i,m}(t)}{\sum_{i=1}^{I} Z_i(t) \check{P}_t(m|i)} \tag{32}$$

$$\bar{\gamma}_{i,m}(t) \equiv \int_{\mathcal{Q}[y]=i} y \check{p}_t(y|i,m) dy \tag{33}$$

with $\check{P}$ and $\check{p}$ from (34) and (35), respectively. In fact the same comments apply here as to the GFMT, as indicated in Appendix A: in the multi-target situation there should be an optimization over the list matching of $\{\gamma_n(t)\}_{n=1}^{M}$ to $\{x_m(t)\}$.

It must be admitted that this is only marginally more satisfying than the original GFMT: it is good that the prior on X is now involved in the filtering, but it is unfortunate that there is no sharing of information between that prior and the refinement of FPA-level data.

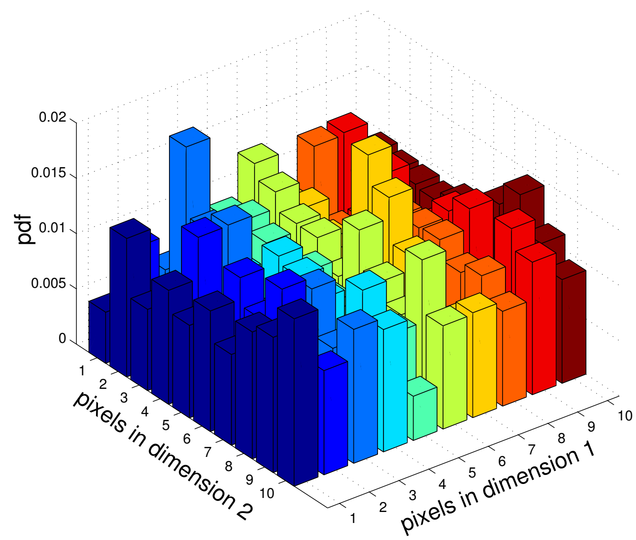

Fig. 3. An example of a pdf for non-uniform clutter, conditioned on the photon being clutter.

C. Introduce Pixel Noise

The model of the GFMT and HPMHT, while adhering to physics in terms of the large count of photons for targets that are not excessively dim, does not match what is usually observed on the FPA. The more common image is more like what is in figure 1. It seems reasonable to model the difference as pixel-level noise. An example is in [6], which uses a Ricean model from [17].

Here we will modify (5) to

$$y_k(t) \sim \left\{ \begin{array}{ll} \mathcal{N}(y; \mathbf{H}x_m(t), \mathbf{R}) & u_k(t) \in \{1, \ldots, M\} \\ f_c(y) & u_k(t) = 0 \end{array} \right. \tag{36}$$

where, as suggested by figure 3, the pdf of a clutter photon is not just a fixed level but is instead piecewise constant. We write

$$f_c(y) = \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2} \frac{v_{i_1,i_2}(t)\mathcal{I}\left(\mathcal{Q}\left[y\right] = (i_1, i_2)\right)}{(\tau_{i_1+1} - \tau_{i_1})(\tau_{i_2+1} - \tau_{i_2})} \tag{37}$$

or using our single-index notation:

$$f_c(y) = \sum_{i=1}^{I} \frac{v_i(t)\mathcal{I}\left(\mathcal{Q}\left[y\right] = i\right)}{(\tau_{i+1} - \tau_i)} \tag{38}$$

We will specify that $\{v_i(t)\}_{i=1}^{I}$ be a set of random variables that are jointly distributed according to $p(V)$ and that sum to unity. An example might be Dirichlet or Jeffreys [3], [11].

$$\check{P}_t(m|i) \equiv Pr(u_k(t)=m|\mathcal{Q}[y_k(t)]=i,\gamma_m(t)) = \begin{cases} \dfrac{\left(\pi_m \int_{y\in i} \mathcal{N}(y;\gamma_m(t),\mathbf{R})dy\right)}{\left(\pi_0 \int_{y\in i} \frac{1}{a_1 a_2} dy\right) + \sum_{l=1}^{M}\left(\pi_l \int_{y\in i} \mathcal{N}(y;\gamma_l(t),\mathbf{R})dy\right)} & m\in\{1,M\} \\ \dfrac{\left(\pi_0 \int_{y\in i} \frac{1}{a_1 a_2} dy\right)}{\left(\pi_0 \int_{y\in i} \frac{1}{a_1 a_2} dy\right) + \sum_{l=1}^{M}\left(\pi_l \int_{y\in i} \mathcal{N}(y;\gamma_l(t),\mathbf{R})dy\right)} & m=0 \end{cases} \tag{34}$$

$$\check{p}_t(y|i,m) \equiv p(y_k(t)|u_k(t)=m,\mathcal{Q}[y_k(t)]=i,\gamma_m(t)) = \mathcal{I}(\mathcal{Q}[y]=i)\times \begin{cases} \dfrac{\mathcal{N}(y;\gamma_m(t),\mathbf{R})}{\int_{y\in i} \mathcal{N}(y;\gamma_m(t),\mathbf{R})dy} & u_k(t)=m\in\{1,M\} \\ \dfrac{\frac{1}{a_1 a_2}}{\int_{y\in i} \frac{1}{a_1 a_2} dy} & u_k(t)=0 \end{cases} \tag{35}$$

$$P_t(m|i) \equiv Pr(u_k(t)=m|\mathcal{Q}[y_k(t)]=i,x_m(t)) = \begin{cases} \dfrac{\left(\pi_m \int_{y\in i} \mathcal{N}(y;\mathbf{H}x_m(t),\mathbf{R})dy\right)}{(\pi_0 v_i(t)) + \sum_{l=1}^{M}\left(\pi_l \int_{y\in i} \mathcal{N}(y;\mathbf{H}x_l(t),\mathbf{R})dy\right)} & m\in\{1,M\} \\ \dfrac{(\pi_0 v_i(t))}{(\pi_0 v_i(t)) + \sum_{l=1}^{M}\left(\pi_l \int_{y\in i} \mathcal{N}(y;\mathbf{H}x_l(t),\mathbf{R})dy\right)} & m=0 \end{cases} \tag{40}$$

Now we have (10) replaced by

$$Q(X,V) = \int \log[p(Z,X,U,V,Y)]p(U,V,Y|Z) \tag{39}$$

in which V is of course the aggregated v_i's. Note especially that (6) does not change, since π_0 refers to the prior probability that a given photon is clutter: if it is, then the pixel to which it is assigned is according to $\{v_i(t)\}_{i=1}^{I}$ in scan t. None of (12)-(19) changes, since these refer to gradients and updates with respect to target positions. One equation that does change is (8), which becomes (40).

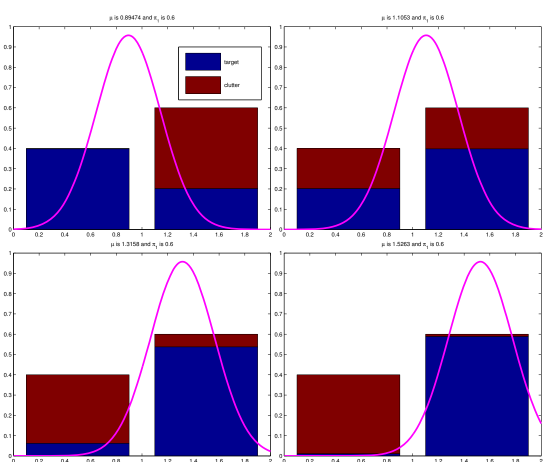

Fig. 4. Illustration of fit of clutter and target for two pixels, observations $Z_i(t) = [0.4\ 0.6]$, and various values of target position μ. Here the point-spread standard deviation is $\sigma = 0.25$

Another change is that while (12) refers to the gradient with respect to target states, we need as well the gradient with respect to the clutter probabilities. We could write (11) as

$$\begin{aligned} Q(X,V) &= \log[p(X)] + \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{1=0}^{M} P_t(m|\mathcal{Q}[y_k(t)]) \\ &\quad \times \int_Y p_t(y_k(t)|\mathcal{Q}[y_k(t)], m)\log[\pi_m p(y_k(t)|x_m(t))] \\ &\quad + \log[p(V)] + \sum_{t=1}^{T}\sum_{k=1}^{K} P_t(0|\mathcal{Q}[y_k(t)])\log[\pi_0 v_{\mathcal{Q}[y_k(t)]}(t)] \end{aligned} \tag{41}$$

From this we get the gradient with respect to V as

$$\nabla Q(X,V) = \frac{\partial \log[p(V)]}{\partial v_i(t)} + \frac{Z_i(t)P_t(0|i)}{v_i(t)} \tag{42}$$

According to Appendix B, we have

$$v_i(t) = \frac{Z_i(t)P_t(0|i)}{\sum_{j=1}^{I} Z_j(t)P_t(0|j)} \tag{43}$$

for every t. It is important to recognize that π_0 must be fixed (and not estimated) if this model is used: the proportion of photons that come from the clutter must be known. If π_0 is estimated an attractive model fit to any FPA data is with $\pi_0 = 1$ and $v_i(t) = \frac{Z_i(t)}{\sum_{j=1}^{I} Z_j(t)}$; this is not useful. Note that (43) is not the only choice of $v_i(t)$; some others, such as the common Gaussian pixel noise assumption, are given in Appendix C.

Now, although (43) can be inserted to EM in (40), it does not get around the problem of $K \to \infty$ in (17): the prior

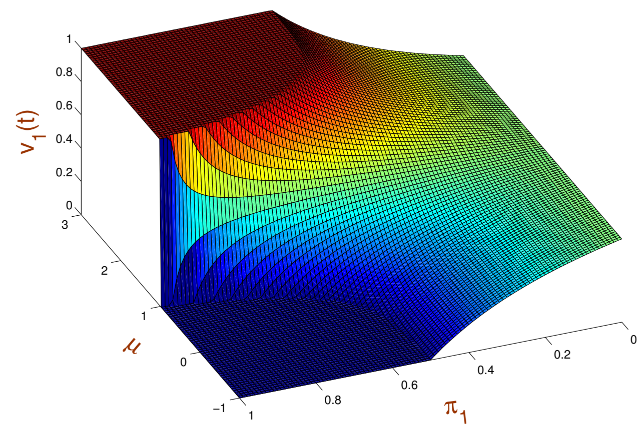

Fig. 5. The manifold of allowable clutter values in the left pixel (#1) as a function of the target position (μ) and the relative target strength (π_1). Here the pixel observations are $Z_i(t) = [0.5\ 0.5]$ and the point-spread standard deviation is $\sigma = 0.5$.

remains effete. So let us write

$$
\begin{aligned}
p(X|Z) &\propto \int p(Z|U,V,Y,X)p(U,V,Y)dUdVdY\,p(X) \\
&= p(X)\int \prod_{t,k}\left[p(z_k(t)|U,V,Y,X)\right] \\
&\qquad \times p(U,V,Y)dUdVdY \qquad (44)
\end{aligned}
$$

Put this way it is clear why the data dominates the prior $p(X)$. What is perhaps less clear is that the "data" term (i.e., everything except $p(X)$) is constant over a manifold. Turning to figure 4 we see a notional case of a single target in two pixels: as the target is moved from left to right the clutter, according to our pixel noise model, adjusts itself to make up the difference between the photon counts in the target and the observations. The case is made more strongly in figure 5, where the manifold of allowable triples $(\mu, \pi_1, v_1(t))$ is shown.

The point is that under this "pixel noise" model there is a manifold of target locations whose likelihoods, represented as the terms in (44) other than $p(X)$, are the same. Within this manifold of points the prior $p(X)$ can have an influence.

VII. Summary

The GFMT introduced a quantum observation model for target tracking based on data that was "histogrammed," as might be the case for focal plane array, beamspace and/or spectral measurements. It was discovered that a detailed model in which the observations were considered as individual quanta (as opposed to aggregated energy levels, for example), could, when coupled with a healthy dose of the EM algorithm, result in a quite simple integrated tracking algorithm that needed no exogenous signal processing.

There was, however, a lacuna in the GFMT's modeling – a small but important gap that the HPMHT cleverly filled. The fix, however, is ad-hoc, and indeed it is not clear whether the approach is the best. Now, what is actually "best" would require a model to underpin it. It is hoped that this paper makes a contribution by suggesting possible models and using the appropriate EM machinery to see what they evince.

Acknowledgment

This research was supported by ONR under contract N000014-13-1-0231 and by NPS NPS via ONR N00244-14-1-0033.

References

[1] Y. Bar-Shalom, P. Willett and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs CT 2011.

[2] Y. Bar-Shalom, X. Rong Li and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley, 2001.

[3] J. Bernardo and A. Smith, *Bayesian Theory*, Wiley, New York, 1994.

[4] D. Crouse, M. Guerriero and P. Willett, "A Critical Look at the PMHT," *Journal on Advances in Information Fusion*, Vol. 4, No. 2, pp. 93-116, December 2009. (D. Crouse, L. Zheng and P. Willett, "Correction to: 'A Critical Look at the PMHT'," *Journal on Advances in Information Fusion*, Vol. 7, No. 1, pp. 97-98, June 2012.)

[5] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.

[6] S. Davey, M. Rutten and B. Cheung, "A Comparison of Detection Performance for Several Track-Before-Detect Algorithms," *ISIF/IEEE Conference on Information Fusion*, Cologne Germany, 2008.

[7] S. Davey and D. Gray, "Integrated Track Maintenance for the PMHT via the Hysteresis Model," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 93-111, January 2007.

[8] S. Davey, M. Wieneke and N. Gordon; "H-PMHT for correlated targets," *Proceeding of Signal and Data Processing of Small Targets (SPIE)*, April 2012.

[9] S. Davey and M. Wieneke and H. Vu, "Histogram-PMHT Unfettered," *IEEE Journal of Selected Topics in Signal Processing,* vol. 7, no. 3, pp.435-447, June 2013.

[10] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1-38, 1977.

[11] L. Campbell, "Averaging Entropy," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 338-339, January 1995.

[12] F. Fleuret, J. Berclaz, R. Lengagne and P. Fua, "Multi-Camera People Tracking with a Probabilistic Occupancy Map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 267-282, February 2008.

[13] T. Luginbuhl and P. Willett, "Tracking a General Frequency Modulated Signal in Noise," *Proceedings of the IEEE Conference on Decision and Control*, Phoenix AZ, December 1999.

[14] T. Luginbuhl and P. Willett, "Estimating The Parameters Of General Frequency Modulated Signals," *IEEE Transactions on Signal Processing*, pp. 117-131, January 2004.

[15] G. McLachlan and P. Jones, "Fitting Mixture Models to Grouped and Truncated Data via the EM Algorithm" *Biometrics*, vol. 44, pp. 571-578, June 1988.

[16] P. Morris, R. Aspden and J. Bell, R. Boyd and M. Padgett, "Imaging with a Small Number of Photons," *Nature Communications*, vol. 6, 2015.

[17] B. Ristic, S. Arulampalam and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.

[18] R. Streit and T. Luginbuhl, "Probabilistic Multi-Hypothesis Tracking", *NUWC-NPT Technical Report 10,428*, February 1995.

[19] R. Streit, "Tracking on Intensity-Modulated Data Streams," *NUWC Technical Report 11,211*, May 2000.

[20] M. Walsh, M. Graham, R. Streit, T. Luginbuhl, and L. Mathews, "Tracking on Intensity-Modulated Sensor Data Streams," *Proceedings of IEEE Aerospace Conference*, vol. 4, pp. 1901-1909, 2001.

[21] M. Wieneke and S. Davey, "Histogram PMHT with Target Extent Estimates Based on Random Matrices," *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pp.1-8, July 2011.
[22] P. Willett, B. Balasingham, D. Dunham and T. Ogle, "Multitarget Tracking Using ABIR Images and the H-PMHT," *SPIE Conference on Signal and Data Processing of Small Targets OPP322*, San Diego CA, August 2013.
[23] P. Willett, Y. Ruan and R. Streit, "The PMHT: Its Problems and Some Solutions," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 738-754, July 2002.

APPENDIX A
MULTI-TARGET GFMT

As indicated in section IV the converged GFMT simply takes its measurements ($\tilde{y}_m(t)$'s) as exact knowledge of a target's location. In the case of a single target the matter is closed; but with multiple targets note that in (44) only $p(X)$ depends on which target is assigned to which measurement: in the limit as $K \to \infty$ the term $\prod_{t=1}^{T} \prod_{k=1}^{K} p(y_k(t)|X)$ would remain the same regardless of permutation of $\{x_m(t)\}$ with respect to m. As such, the GFMT does not, in the multi-target case, completely ignore the prior $p(X)$: the best set of tracks would be from solving

$$\min_{\{n(t)\}} \left\{ p(\{x_{n(t)}\}|\{\tilde{y}_m(t)\}) \right\} \tag{45}$$

meaning a multi-list assignment of targets $\{x_m(t)\}$ to (noise-free) measurements $\{\tilde{y}_m(t)\}$. This is a small consolation; but does demonstrate that even in the GFMT the prior is not always completely irrelevant.

APPENDIX B
THE EFFECT OF PIXEL NOISE: DERIVATION OF (43)

Consider the Dirichlet distribution

$$p(\{v_i(t)\}) = \frac{\Gamma(\nu I)}{\Gamma(\nu)^I} \prod_{i=1}^{I} v_i(t)^{\nu-1} \tag{46}$$

over the finite hyperplane $\sum_{i=1}^{I} v_i(t) = 1,\ v_i(t) \geq 0,\ \forall i$. We see that (43) is exactly true for any K in the case of the "canonical" Dirichlet case $\nu = 1$, meaning the v's are uniform over their domain. It is also exactly true for the exponential case that $p(v_i(t)) = \beta e^{-\beta v_i(t)}$ on the same finite hyperplane.

In fact, turning to (42), we see that *in general* the first term (involving the prior on V) does not depend on the number of photons K unless such dependence is forced by the modeler. As such, the gradient evaporates in the limit as $K \to \infty$, and the only remaining term is the Lagrange multiplier λ enforcing that the that v's sum to unity. Equation (43) follows.

APPENDIX C
THE EFFECT OF PIXEL NOISE: ALTERNATIVES TO (43)

Appendix B mentioned that one might "force" dependence on the number of photons. Consider the case of (46) with $\nu = \nu_0 K$: with large K this is a form of the Dirichlet that strongly encourages small v's. Solving (42) we get

$$v_i(t) = \frac{[Z_i(t)P_t(0|i) - \nu]^+}{\sum_{j=1}^{I} [Z_j(t)P_t(0|j) - \nu]^+} \tag{47}$$

in the case $K \to \infty$, and in which the notation $[\xi]^+ = \xi$ if $\xi \geq 0$ and zero otherwise.

Consider also the case that $p(V)$ is such that $v_i(t)$ is Gaussian with mean μ and variance σ^2; and that $\sum_{i=1}^{I} v_i(t) = 1$, $v_i(t) \geq 0,\ \forall i$ as discussed earlier. For this rather common "Gaussian pixel noise" assumption we get

$$v_i(t) = \left[\frac{\mu + \lambda}{2} + \sqrt{\left(\frac{\mu+\lambda}{2}\right)^2 + \sigma^2 Z_i(t) P_t(0|i)} \right]^+ \tag{48}$$

and λ is chosen to make the $\sum_{i=1}^{I} v_i(t) = 1$.

It is not apparent that either (47) or (48) is a good model. What perhaps should be stressed, however, is that neither assumption is much different in spirit from (24) to amplify the prior information.