HAL
open science

Adaptive birth for the GLMB filter for object tracking in satellite videos

Camilo Aguilar, Mathias Ortner, Josiane Zerubia

To cite this version:

Camilo Aguilar, Mathias Ortner, Josiane Zerubia. Adaptive birth for the GLMB filter for object tracking in satellite videos. MLSP 2022 - IEEE International workshop on machine learning for signal processing, Aug 2022, Xi'an, China. 10.1109/MLSP55214.2022.9943411 . hal-03715704

HAL Id: hal-03715704

https://inria.hal.science/hal-03715704v1

Submitted on 6 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE BIRTH FOR THE GLMB FILTER FOR OBJECT TRACKING IN SATELLITE VIDEOS

Camilo Aguilar[⋆] *Mathias Ortner*[†] *Josiane Zerubia*[⋆]

[⋆] Inria, Université Côte d'Azur, Sophia Antipolis, France
[†] Airbus Defense and Space, Toulouse, France

ABSTRACT

The Generalized Labeled Multi-Bernoulli (GLMB) filter attains remarkable results in Multi-Object Tracking (MOT). Nevertheless, the GLMB filter relies on strong assumptions such as prior knowledge of targets' initial state. Pragmatic scenarios such as satellite video object tracking challenge these assumptions as objects appear at random locations and object detectors output numerous false positives. We present an enhanced version of the GLMB filter that learns from previous trajectories to estimate accurate hypotheses initializations. We keep track of previous target states and use this information to sample the initial velocities of new-born targets. This addition significantly improves the performance of the GLMB in videos with low Frames Per Second (FPS), where the target's initial states are paramount for object tracking. We test this enhanced GLMB filter versus comparable trackers and previous solutions for the GLMB filter and show that our filter obtains better performance. Code is available at https://github.com/Ayana-Inria/GLMB-adaptive-birth-satellite-videos

***Index Terms*—** GLMB, RFS, adaptive birth, multitarget tracking, remote sensing

1. INTRODUCTION

Novel earth observation satellites and Unmanned Aerial Vehicles (UAV) capture content-rich ground information that opens opportunities for emerging applications. Many of these applications rely on detecting and tracking targets from high-resolution videos. For example, the Jilin1-1 satellite constellations contribute to Chinese urban investigation by tracking vehicles and driver behavior across urban areas [1]. Similarly, the French Zephyr Unmanned Aerial System (UAS) contributes to maritime surveillance by monitoring ships for piracy or smuggling of goods and narcotics [2], or the American Maxar Worldview's satellite constellation contributes to military assessments in regions without local sensors [3].

MOT frameworks normally rely on data association such as Joint Probabilistic Data Association (JDPA) [4], Multiple Hypotheses Tracking (MHT) [5] or Random Finite Sets (RFS) [6]. Among these approaches, the RFS framework has gained popularity due to its elegant Bayesian formulation to jointly model object dynamics (births and transitions) and measurements as set-valued random variables. Common RFS-based filters include the Probability Hypothesis Density (PHD) filter [7] or the Cardinalized PHD (CPHD) filter [8]. These filters provide robust and computationally efficient solutions using first-order approximations for the multi-object distribution; however, they fail to assign object labels and recover object trajectories.

Recently, Vo and Vo [9] proposed the labeled RFS framework to model both multi-object states and multi-object trajectories based on a rigorous mathematical foundation. The most popular approach is the Generalized Labeled Multi-Bernoulli (GLMB) filter and its computationally efficient derivation [10]. The GLMB filter has provided a solution for MOT requirements across several fields, such as in biological images [11] or radar applications [12].

However, despite the solid theoretical background, the GLMB filter highly depends on strong assumptions such as prior knowledge of the birth components' initial location and known clutter rate. These assumptions do not always hold on real-world scenarios as targets appear and disappear at often random locations. For example, in remote sensing images, ground vehicles can appear from incoming roads or parking lots; planes can appear from airports, and pedestrians can appear from buildings. Additionally, targets in satellite images are often tiny and have a white-blob-looking appearance. Targets are frequently lost in clutter, and object detectors often output several clutter measurements from static white-blob-looking objects.

To compensate for the object initialization drawback, numerous works have proposed the "measurement-driven birth" [13], which consists in initializing components at unassigned, low-likelihood measurements. This approach presents two main problems: the initialization does not discriminate between clutter and birth measurements; hence, it can initialize tracks at faulty locations. Second, the velocity component is unknown, which causes the filter to lose fast-moving targets in future frames.

In this paper, we propose a birth initialization method

based both on low likelihood components and on component history. We exploit the fact that ground targets in satellite images often follow similar directions to previously tracked objects (normally dictated by the direction of highways and roads). We utilize this information to adaptively learn the initial state of the birth component measurements and hence, improve the filter performance for fast-moving targets in low FPS datasets. We perform tests in a challenging, low-resolution, low-FPS dataset and show that our filter outperforms other state-of-the-art methods and variations of RFS filters.

We present further details of relevant literature review in Section 2, we discuss about the random finite set background in Section 3, we detail the proposed method in Section 4, we show results in Section 5, and finally we discuss our main findings and future work in Section 6.

2. RELATED WORK

The classical measurement-driven birth constitutes the conventional tool to approximate birth intensities in RFS-based filters. For example, [13, 14] use measurements with low association likelihood from the previous frame to initialize new components. This approach misses the first few detections of a new target but is robust to clutter measurements. Similarly, [15] uses measurements from the current frame to drive the birth intensity. This initialization captures targets' initial measurements but suffers from false positives initialized at clutter detections. These techniques usually initialize hypotheses with random velocity components and large uncertainties. The hypothesis velocity is updated during the filter update step if the filter associates the birth component with the object's subsequent measurement; otherwise, the filter initializes a new hypothesis with a new label and the original track is lost.

Additionally, the work in [16] uses a pre-processing step to approximate the birth distribution. This approach accounts for three measurements before estimating the birth RFS, and hence, it has a rough initial approximation of the object's initial conditions. This methodology helps to reduce clutter detections and to estimate the initial velocity at the cost of missing the first three detections for each object. Furthermore, the "birth" object proposal step highly depends on the association algorithm used as a pre-processing step.

Several papers present adaptive clutter rate estimation by combining the CPHD with the GLMB [8, 17]. These approaches consist of using the CPHD filter's cardinality estimation to infer the detection and clutter rates of the GLMB. While these methods have shown improved performance with respect to the original GLMB filter, they still rely on pre-hard-coded parameters and do not exploit past track information.

Fu et al. [18] proposed an online dictionary learning method to learn object appearance and discriminate target measurements from clutter measurements based on the correlated dictionary features. This approach shows promise; however, satellite image targets often lack distinctive features to discriminate them from clutter measurements.

Finally, it is worth mentioning deep-learning approaches such as MT3 [19]. This work uses transformers to track objects in highly cluttered and non-linear scenarios with a track-by-detection paradigm. MT3 obtained comparable results to model-based filters at the cost of using 18 million parameters and several days to train the network. However, target dynamics in satellite images can be modeled well with linear motion and matrix operations using significantly fewer parameters.

3. BACKGROUND

We model a single target kinematic state as $x = [p, v]^T$ where $p \in \mathbb{R}^2$ denotes the target 2D location coordinates and $v \in \mathbb{R}^2$ denotes the target 2D velocity components. Similarly, we model a single measurement as $z = [\bar{p}]$, where $\bar{p} \in \mathbb{R}^2$ represents the 2D coordinates for each measurement.

3.1. Labeled RFS

The RFS framework models multi-object states and measurements as finite-set value random variables. In this work, we use the labeled version of the traditional RFS framework, which assigns a unique label to each element in the multi-object RFS. The addition of labels allows the filter to recover both object states and object trajectories.

Let $\mathbb{X}$ be the state space and $\mathbb{L}$ be a discrete label space. We model a single labeled target as $\mathbf{x} = (x, \ell)$, where $x \in \mathbb{X}$ is the target kinematic state and $\ell \in \mathbb{L}$ is the label associated to the target. Subsequently, we characterize the collection of labeled multi-target states at frame k as $\mathbf{X}_k = \{\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}\} \in \mathcal{F}(\mathbb{X})$, where $\mathcal{F}(\cdot)$ is a collection of finite subsets and $n_k \in \mathbb{N}$ is a random variable associated with the number of elements in $\mathbf{X}_k$. Similarly, we define the measurement space as $\mathbb{Z}$ and define the measurement RFS as $Z_k = \{z_1^{(k)}, z_2^{(k)}, \ldots, z_{m_k}^{(k)}\} \in \mathcal{F}(\mathbb{Z})$, where m_k is a random variable denoting the number of measurements and $z_j^{(k)}$ is the jth measurement at frame k.

Following a Bayesian setup, the multitarget density $\pi_k(\mathbf{X}_k)$ is conditioned on measurements and is propagated through time following a prediction, $\pi_{k+1}(\mathbf{X}_{k+1})$ and update step $\pi(\mathbf{X}_{k+1}|Z_{k+1})$.

3.2. The GLMB Filter

The GLMB filter provides an analytical solution to the Bayes multi-target tracking by propagating a labeled multi-Bernoulli (LMB) density in time. For its implementation, the GLMB filter uses its analytical form known as δ-GLMB [10] given by the equation:

$$\pi(\mathbf{X}) = \Delta(\mathbf{X}) \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi} \omega^{(I,\xi)} \delta_I(\mathcal{L}(\mathbf{X}))[p^{\xi}]^{\mathbf{X}} \quad (1)$$

Where $I \in \mathcal{F}(\mathbb{L})$ denotes a set of object labels, $\xi \in \Xi$ denotes the history of measurement-target association maps, and $\omega^{(I,\xi)}$ is a non-negative weight for each (I, ξ) hypothesis. The term $[p^{\xi}]^{\mathbf{X}}$ refers to the exponential notation of the product $\prod_{x \in \mathbf{X}} p^{\xi}(x)$ with $[p^{\xi}]^{\emptyset} = 1$ and the term $p^{\xi}(\cdot)$ denotes a probability density describing a multi-Bernoulli component's kinematic state. $\mathcal{L}$ denotes the labeling projecting function $\mathcal{L} : \mathbb{X} \times \mathbb{L} \rightarrow \mathbb{L}$ and $\Delta(\mathbf{X})$ denotes the distinct label operator that is non-zero only when the labels in $\mathbf{X}$ are unique. Throughout this paper, we use the bold notation $\mathbf{x}$, $\mathbf{X}$ for labeled states and the regular notation x, X for non-labeled states, i.e. $\mathbf{x} = (x, \ell)$

The GLMB filter is propagated in time with the prediction step and update step. For this work, we use the joint-prediction and update proposed by Vo and Vo in [10]. This efficient implementation uses Gibbs sampling to estimate likely realizations of the filter at each time step without exhaustively computing every possible combination of predicted and updated hypotheses. Similarly, we perform inference and cardinality estimation following the procedure listed by Vo and Vo [10].

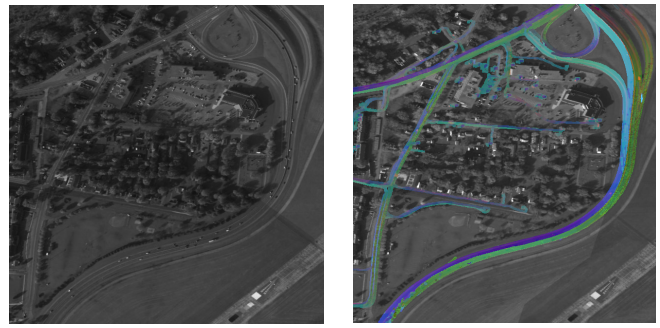
4. ADAPTIVE MEASUREMENT DRIVEN BIRTH

4.1. Learning the Initial Velocities

During each frame, we perform inference to obtain the estimated label object state set $\hat{\mathbf{X}}_k = \{\hat{\mathbf{x}}_1^{(k)}, \hat{\mathbf{x}}_2^{(k)}, \ldots \hat{\mathbf{x}}_{n_k}^{(k)}\}$. Each estimated object $\hat{\mathbf{x}}_i^{(k)}$ contains an assigned label $\hat{\ell}_i$ and an estimated kinematic state $\hat{x}_i^{(k)} = [\hat{p}_i, \hat{v}_i]^T$, where $\hat{p}_i$ represents the estimated target location and $\hat{v}_i$ denotes the estimated target's velocity at frame k, and $i = \{1, 2, \cdots, n_k\}$ denotes the object index. Every tracked element $\hat{\mathbf{x}}_i^{(k)}$ leaves a trace of its state at every inferred frame. Therefore, we create a velocity history map $h_v^k : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, which updates iteratively at every frame k and $h_v^0(i,j) = [0,0]^T \, \forall \, i, j \in \mathbb{N}$. For the state $\hat{x}_i^{(k)}$ of each inferred object $\hat{\mathbf{x}}_i^{(k)} \in \hat{\mathbf{X}}_k$, we learn the velocity history map h_v^k as follows:

$$h_v^k(\hat{p}_i) = \begin{cases} \frac{h_v^{k-1}(\hat{p}_i) + \hat{v}_i}{2} & \text{if} \quad ||h_v^{k-1}(\hat{p}_i)|| \neq 0 \\ \hat{v}_i & \text{if} \quad ||h_v^{k-1}(\hat{p}_i)|| = 0 \end{cases} \tag{2}$$

Where $\hat{p}_i$ denotes the ith object estimated location, and $\hat{v}_i$ denotes the estimated velocity at frame k. The equation updates the velocity history map by averaging with new inferred velocity values. Additionally, we perform a linear interpolation between each labelled object in $\hat{\mathbf{X}}_k$ with its past state in $\hat{\mathbf{X}}_{k-1}$ (in the case a target with the same label exists both in frames). Fig. 1 shows a sample velocity history map created from several time instances. The velocity 2D vector is encoded in the color channels.



(a) Sample image (b) Learned velocities $h_v^k(p)$

Fig. 1. Sample learned directions from past vehicles.

4.2. Adaptive Measurement Driven Birth Intensity

During each iteration, we split new measurements into surviving and birth measurements $Z_k = Z_k^S \cup Z_k^B$ depending on their distance to the predicted GLMB components. We call a measurement "surviving measurement" if is located less than a T_{birth} distance from the location coordinates of the nearest Bernoulli component. Sequentially, we define the density of new-born targets $\pi_B(\cdot)$ as:

$$\pi_B(\mathbf{X}_+) = \delta(\mathbf{X}_+)\omega_B(\mathcal{L}(\mathbf{X}_+))[p_B]^{\mathbf{X}_+} \tag{3}$$

Where $\mathbf{X}_+$ denotes the predicted labeled object states, ω_B is the probability of a birth component, and p_B denotes the probability distribution of the kinematic states of the new-born targets. We define p_B as:

$$p_B(x, \ell; z) = \frac{1}{|Z_k^B|} \sum_{z_j \in Z_k^B} \mathcal{N}(x; m_B(z_j), P_B) \tag{4}$$

Where P_B is the initial covariance hyperparameter, $\ell \in \mathbb{N}$ is an assigned new label, and the function $m_B(\cdot)$ maps the birth measurements $z_j \in Z_k^B$ to the kinematic target states. The function $m_B(\cdot)$ has a major impact in the filter's performance and is defined as:

$$m_B(z_j) = [\bar{p}_j, h_v^k(\bar{p}_j)]^T \tag{5}$$

where $\bar{p}_j$ denotes the location of measurement z_j, and $h_v^k(\bar{p}_j)$ is the learned velocity function evaluated at location $\bar{p}_j$. This method differs from the classical "measurement driven birth" [13] where the velocity component of $m_B(\cdot)$ is unknown, random, or initialized to zeros.

5. EXPERIMENTS

5.1. Metrics

We evaluate the experiments using the ClearMOT [20] metric. This metric reports the Multiple Objects Tracking Precision ($MOTP$) and the Multiple Object Tracking Accuracy

($MOTA$). The $MOTP$ depends on the average distance error between the track hypothesis and ground truth trajectories and ranges in $[0, \infty)$, where 0 denotes a perfect score. The $MOTA$ depends on False Positives (FPs), False Negatives (FNs), Identity Switches ($IDSs$) and the Number of Ground Truth objects ($NGTs$), and its score ranges from $(-\infty, 1]$, where 1 denotes a perfect score. Throughout this work, we call a hypothesis True Positive (TP) if it has a uniquely matched ground truth object less than $d = 5$ pixels away.

Additionally, we name a trajectory Mostly Tracked (MT) if more than 80% of its ground truth trajectory is tracked. Similarly, we say a trajectory is Partially Tracked (PT) if it is not MT, but more than 20% of its trajectory is tracked; otherwise, we classify the trajectory as Mostly Lost (ML).

We test our method and compare it with the results obtained using the classical "measurement-driven" birth estimation (as used in [14, 15, 17]). We also compare with the modified track-GMPHD to recover trajectories proposed in [21] and with the SORT [22] algorithm as it has become a baseline for MOT-based works.

5.2. Results

We test our method with the WPAFB 2009 dataset available online [23]. This dataset contains videos recorded by a single sensor of the Wright Peterson Airforce Base in Ohio, USA. The images span an area of 19 km^2 at a resolution of 0.25m/pixel and a frame rate of 1.25 HZ over 512 frames. We use a downsampled version of this dataset at a resolution of 1.0m/pixel to mimic satellite videos and we register the images using the ORB stabilization algorithm similar to [21]. We evaluate tracking results in specific areas of interest comprised of high-speed highways and a crossing. Particularly, we choose the areas of interest (AoIs) 02, 34, and 42 as shown in [24] to be consistent with the literature. We choose these regions as they present fundamental challenges for object trackers: objects move at high speeds, objects are often in crowded environments, and trajectories intersect regularly.

Table 1 shows that adding our adaptive birth improves the GLMB filter scores of the trajectory quality metrics across all three AoIs. For example, our filter attains a mean of 76.35% of MT trajectories while the traditional GLMB attains 53.59%, and the T-GMPHD attains only 20.75% of the MT trajectories. The reason behind this significant improvement is the adaptive birth estimation. Our approach initializes hypotheses with accurate velocity components. Therefore, the filter performs a correct data association between the predicted birth hypothesis and the actual frame measurements. The filters implemented with the traditional "measurement-driven birth" lose targets as the predicted birth hypotheses fail to track the target's future measurements. If the filters fail to follow targets, the hypotheses are considered clutter, or a new birth component is initialized at future untracked measurements. This issue is also manifested in the large number of identity switches shown in the seventh column of Table 1 where our approach obtains the lowest number of IDSs with respect to the number of detected objects.

The GLMB with default birth estimations obtains the lowest mostly lost (ML) scores. For all three cases, it has an ML score of 0.58%; this means the filter partially tracks objects but cannot fully recover whole trajectories. Similarly, the SORT tracker obtains the least number of IDSs across all three AoIs as it fails to track a large majority of objects. It is worth noting that in this work, we also consider static objects; hence, the results are artificially inflated even for weaker trackers such as SORT and the track-GMPHD.

6. CONCLUSION AND FUTURE WORK

This paper presented an adaptive birth estimation for the GLMB filter tailored for tracking objects in satellite images. The adaptive birth enhances the classical measurement-driven birth approach with the addition of velocity component initialization. Our approach significantly improves fast-moving object tracking on high-speed roads; hence, it reduces label switches, the number of false positives and false negatives.

While this addition shows significant advantages for tracking objects on high-speed roads, Table 1 shows that our tracker still has room for improvement. For example, visual appearance information could improve the target-measurement association and reduce clutter measurements. Similarly, the inclusion of visual information in different frames could contribute to estimating targets' initial velocities.

7. ACKNOWLEDGMENT

The authors are grateful to the OPAL infrastructure from Université Côte d'Azur for providing computing resources and support. Additionally, the authors would like to thank BPI France for the financial support under the LiChiE contract.

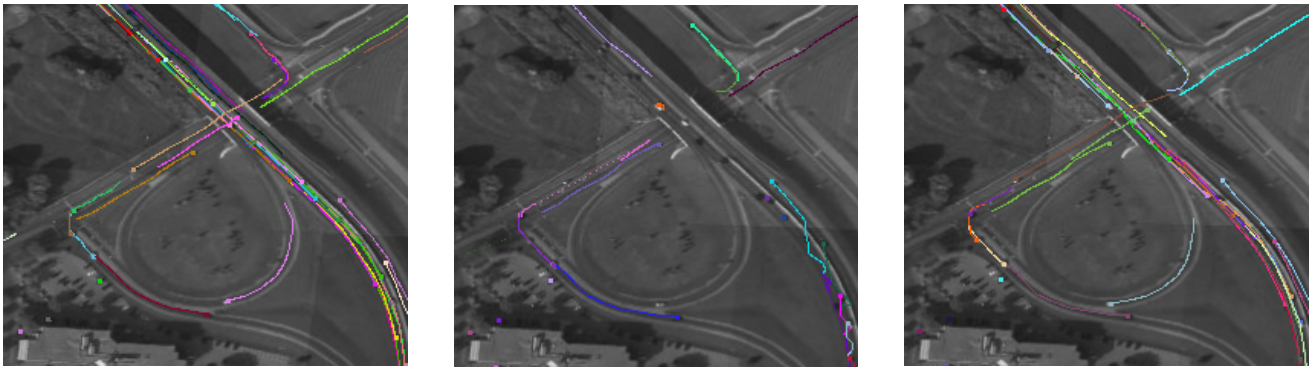
8. REFERENCES

[1] Bo Du, Yujia Sun, Shihan Cai, Chen Wu, and Qian Du, "Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 168–172, 2018.

[2] Sanja Bauk, Nexhat Kapidani, Žarko Lukšić, Filipe Rodrigues, and Luís Sousa, "Review of unmanned aerial systems for the use as maritime surveillance assets," in *24th IEEE International Conference on Information Technology (IT)*, 2020, pp. 1–5.

Area	**Method**	GT	MT↑	PT↑	ML↓	FP ↓	FN ↓	IDS↓	MOTA↑	MOTP↓
02	**Proposed**	654	**477**	165	12	116	**3728**	1795	**0.871**	**0.523**
	GLMB	654	321	**331**	**2**	240	6760	4244	0.742	0.547
	T-GMPHD	654	100	284	270	1706	19172	2478	0.464	0.554
	SORT	654	130	49	475	**93**	22221	**325**	0.478	0.824
34	**Proposed**	690	**517**	166	7	**173**	**4860**	**2606**	**0.896**	**0.375**
	GLMB	690	374	**311**	**5**	408	9006	6483	0.783	0.400
	T-GMPHD	690	139	247	304	3076	26603	2659	0.559	0.520
	SORT	690	66	17	607	555	59253	191	0.173	0.838
42	**Proposed**	287	**233**	50	4	**23**	**1375**	526	**0.905**	**0.382**
	GLMB	287	165	**120**	**2**	64	2066	1297	0.831	0.428
	T-GMPHD	287	77	94	116	887	6669	463	0.604	0.422
	SORT	287	76	24	187	0	8337	**197**	0.576	0.863
Average (%)	**Proposed**	-	**76.35**%	22.24%	1.41%	-	-	-	**0.891**	**0.426**
	GLMB	-	53.59%	**45.83**%	**0.58**%	-	-	-	0.785	0.458
	T-GMPHD	-	20.75%	37.32%	41.92%	-	-	-	0.542	0.498
	SORT	-	18.64%	6.11%	75.25%	-	-	-	0.409	0.841

Table 1. Tracking scores. The arrow's direction represents better scores.



(a) Ground truth trajectories (b) GLMB + measurement driven birth (c) GLMB + proposed adaptive birth

Fig. 2. AoI 2 sample tracking results.

[3] Dexter Jagula, "A boom with a view: The satellite-imaging industry is exploding. Here's how to take advantage of it," *IEEE Spectrum*, vol. 59, no. 3, pp. 38–43, 2022.

[4] Thomas E. Fortmann, Yaakov Bar-Shalom, and Molly Scheffe, "Multi-target tracking using joint probabilistic data association," in *19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, 1980, pp. 807–812.

[5] Donald Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

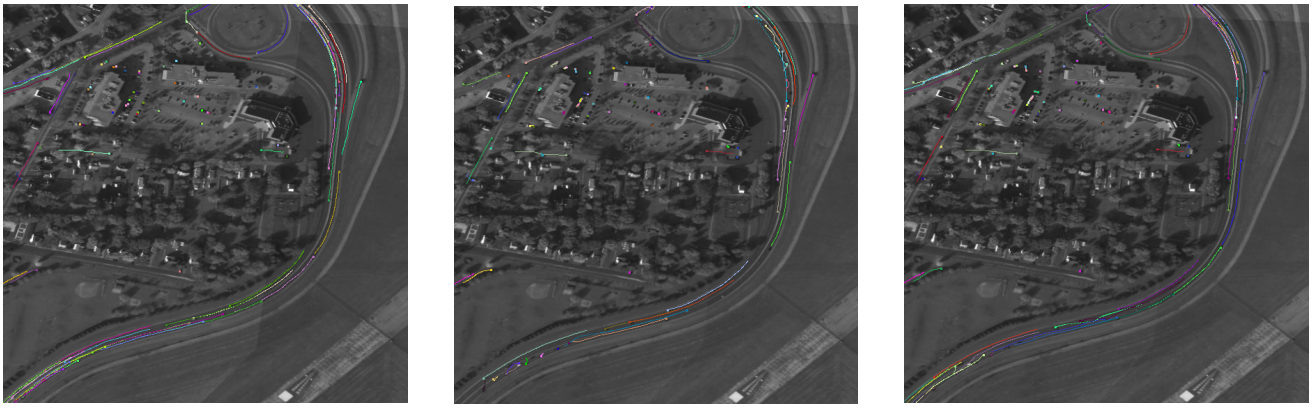
[6] Ronald P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Artech House, Inc., USA, 2007.

[7] Ronald P.S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, pp. 1152 – 1178, 2003.

[8] Ba-tuong Vo, Ba-ngu Vo, and Antonio Cantoni, "The cardinalized probability hypothesis density filter for linear Gaussian multi-target models," in *IEEE 40th Annual Conference on Information Sciences and Systems*, 2006, pp. 681–686.

[9] Ba-Tuong Vo and Ba-Ngu Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, pp. 3460–3475, 2013.

[10] Ba-Ngu Vo, Ba-Tuong Vo, and Hung Gia Hoang, "An efficient implementation of the generalized labeled



(a) Ground truth trajectories (b) GLMB + measurement driven birth (c) GLMB + proposed adaptive birth

Fig. 3. Area 34 tracking results.

multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1975–1987, 2017.

[11] Michael Beard, Ba Tuong Vo, and Ba-Ngu Vo, "A solution for large-scale multi-object tracking," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2754–2769, 2020.

[12] Weihua Wu, Hemin Sun, Yichao Cai, Surong Jiang, and Jiajun Xiong, "Tracking multiple maneuvering targets hidden in the DBZ based on the MM-GLMB filter," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2912–2924, 2020.

[13] Shoufeng Lin, Ba Tuong Vo, and Sven E. Nordholm, "Measurement driven birth model for the generalized labeled multi-Bernoulli filter," in *IEEE International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2016, pp. 94–99.

[14] Branko Ristic, Daniel Clark, Ba-Ngu Vo, and Ba-Tuong Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1656–1668, 2012.

[15] Michele Pace, *Stochastic models and methods for multi-object tracking*, PhD thesis, Université Sciences et Technologies - Bordeaux, I, France, July 2011.

[16] Zong-Xiang Liu, Jie Gan, Jin-Song Li, and Mian Wu, "Adaptive generalized labeled multi-Bernoulli filter for multi-object detection and tracking," *IEEE Access*, vol. 9, pp. 2100–2109, 2021.

[17] Stephan Reuter, Daniel Meissner, Benjamin Wilking, and Klaus Dietmayer, "Cardinality balanced multi-target multi-Bernoulli filtering using adaptive birth distributions," in *Proceedings of the IEEE 16th International Conference on Information Fusion*, 2013, pp. 1608–1615.

[18] Zeyu Fu, Pengming Feng, Federico Angelini, Jonathon Chambers, and Syed Mohsen Naqvi, "Particle PHD filter based multiple human tracking using online group-structured dictionary learning," *IEEE Access*, vol. 6, pp. 14764–14778, 2018.

[19] Juliano Pinto, Georg Hess, William Ljungbergh, Yuxuan Xia, Lennart Svensson, and Henk Wymeersch, "Next generation multitarget trackers: Random finite set methods vs transformer-based deep learning," in *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8.

[20] Keni Bernardin and Rainer Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1–10, 2008.

[21] Camilo Aguilar, Mathias Ortner, and Josiane Zerubia, "Small moving target MOT tracking with GM-PHD filter and attention-based CNN," in *31st IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2021, pp. 1–6.

[22] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft, "Simple online and realtime tracking," in *IEEE International Conference on Image Processing (ICIP)*, 2016.

[23] U.S. Air Force Research Laboratory, "Wright-Patterson air force base (WPAFB) dataset," 2009, data retrieved from the SDMS website, https://www.sdms.afrl.af.mil/index.php.

[24] Rodney LaLonde, Dong Zhang, and Mubarak Shah, "Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.