IAC–25–A6.9.5.x99102

Probabilistic Tracklet Clustering and Initial Orbit Determination

Yihang Jiang
Beijing Institute of Technology, Beijing 100081, China, yihangjiang@126.com

Han Cai
Beijing Institute of Technology, Beijing 100081, China, caihanspace@gmail.com

Yang Yang
School of Mechanical and Manufacturing Engineering, UNSW, Sydney, NSW 2052, Australia, yang.yang16@unsw.edu.au

Jingrui Zhang
Beijing Institute of Technology, Beijing 100081, China, zhangjingrui@bit.edu.cn

Bing Ju
Beijing Aerospace Control Center, Beijing 100094, China, jucice@126.com

Ying Liu
Beijing Aerospace Control Center, Beijing 100094, China, yingliubc@163.com

Space Situational Awareness (SSA) is critical for monitoring the increasing number of space objects. A key component of SSA is tracklet clustering, which involves integrating fragmented observation data from multiple sensors to construct continuous trajectories. Effective tracklet clustering facilitates object identification and initial orbit determination (IOD), thereby enhancing collision avoidance and space traffic management. The primary challenge in tracklet clustering is managing the vast volume of observation data. Deploying mega-constellations has significantly expanded the number of observable space objects. Traditional approaches struggle to establish accurate correlations within such large datasets, particularly in the case of constellation satellites. Moreover, optical tracklets frequently arise from short observation arcs, leading to significant uncertainty in angular velocity, which further complicates the association of multiple tracklets with a single object.

To overcome these challenges, this paper develops a probabilistic tracklet clustering method and a robust IOD method based on clustering. The process begins with an initial association executed via the Boundary Value Problem-Constrained Admissible Region Optimization method (BVP-CAR-Opt). The association probability between two tracklets is assessed through the Mahalanobis distance between the measured and estimated angular-rate information. The tracklet clustering is formulated as an optimization problem. The optimization objective is to maximize the likelihood of the association matrix, which is calculated based on the association probabilities obtained through the BVP-CAR-Opt method. This progress achieves efficient and accurate tracklet clustering by solving the optimal association matrix, avoiding the exhaustive enumeration of all hypotheses. Once the clusters are determined, the weighted hypothetical orbits are generated by the BVP-CAR-Opt according to their corresponding association probabilities. These weighted hypothetical orbits are subsequently utilized in the least squares algorithm to yield the initial orbit. Incorporating weighting factors significantly reduces the relative influence of the assumed orbits on error estimation, thereby ensuring a stable and consistent initial orbit.

To evaluate the performance of the developed method, one simulation scenario was employed and compared with Markov clustering techniques. Optical observation data were generated from a LEO space-based sensor. The tracklet clustering was applied to 30 randomly selected space objects from the 18th Space Defense Squadron public catalog. The new approach demonstrates a higher tracklet clustering accuracy compared with Markov clustering, affirming its effectiveness in typical scenarios.

I. Introduction

Space Situational Awareness (SSA) plays a crucial role in ensuring the safe and sustainable use of space, especially as the number of Resident Space Objects (RSOs) continues to increase rapidly. Maintaining an accurate and up-to-date catalog of Resident Space Objects (RSOs) is a cornerstone of SSA, enabling conjunction analysis, collision avoidance, and long-term sustainability of space operations. However, the proliferation of mega-constellations has dramatically increased the density of objects in orbit,[1] creating new challenges in catalog maintenance.

In SSA, both ground-based and space-based optical sensors are widely used due to their high angular resolution and wide coverage. However, because of observational constraints such as lighting conditions[2] and sensor-target dynamics,[3] they typically produce short-arc observations referred to as tracklets. A single tracklet is typically too short to support reliable Initial Orbit Determination (IOD).[4] Therefore, cataloging relies on the effective association of multiple tracklets that originate from the same object.

The cataloging workflow typically includes two key phases: Tracklet-to-Orbit Association (T2OA) and Tracklet-to-Tracklet Association (T2TA), followed by tracklet clustering. In T2OA, newly acquired tracklets are compared against known cataloged orbits using propagation and statistical matching.[5] Unmatched tracklets are classified as Uncorrelated Tracklets (UCTs) and require further analysis. T2TA, also known as tracklet association, attempts to match UCTs with each other by evaluating whether they may correspond to the same RSO. Once sufficient pairwise associations are established, tracklet clustering is employed to group consistent tracklets into sets that represent individual RSOs.[6]

Tracklet association and clustering thus form the backbone of the uncataloged object processing pipeline. They are critical to enabling reliable IOD and the subsequent creation of new catalog entries. In response to this need, various tracklet association and tracklet clustering methods have been proposed. Recently, tracklet association methods can be roughly categorized into two types: the Initial Value Problem (IVP) and the Boundary Value Problem (BVP) method. The IVP method estimates an orbit by assuming initial distance and velocity at the first observation epoch and propagating forward, while the BVP method fits an orbit by connecting two observations through Lambert's solution using assumed distances at both epochs. These two methods, which transform tracklet association into an optimization problem, were proposed by Siminski et al.[7] Then, Ref.[8] improved the IVP method by modifying the loss function in a nonsingular canonical space, which makes the optimization process smoother and more efficient. The Ref.[6] based on the IVP method, utilizes the bond energy algorithm for clustering, and introduces a splitting algorithm to reduce computational load and handle erroneous associations. Despite their computational efficiency, the IVP method is prone to local optima, and the solution range of the BVP method is extensive, thereby increasing the misassociation rates of the two methods. To address this, the BVP-Constrained Admissible Region (CAR) method was developed in references,[9–11] though related to BVP, it is not strictly an optimization method. Instead, it incorporates additional constraints under the assumption of Lambert orbital dynamics to filter and assess tracklets for association. Additionally, there are some other tracklet association methods based on optimization. Pirovano et al.[12] decomposed the six-dimensional association problem into two-dimensional and four-dimensional optimization problems. The uncertainty and nonlinear dynamics of short-arc observations can be managed through differential algebra and automatic domain splitting techniques. The Ref.[13] links two tracklets by finding the minimum-energy control policy between them through optimization and iteration. This method is applicable for both maneuvering and non-maneuvering spacecraft. However, it is limited to distinguishing between closely spaced objects.

In tracklet clustering research, the Markov Clustering (MCL) algorithm, originally proposed for graph clustering by Van Dongen,[14] has also been applied to the tracklet association problem. Yanez et al.[15] integrated an association method with the MCL algorithm to identify new space objects. By simulating random walks on a graph, MCL effectively differentiates between true and false associations. Nevertheless, the inflation parameter is user-defined and sensitive, which can result in variable clustering outcomes, and the need for further evaluation of its scalability to large datasets. Zhu et al.[16] modeled tracklets as data points, orbits as cluster centers, and residuals as distance metrics, applying the mean-shift clustering algorithm for tracklet association. However, this approach exhibits a high false-positive rate when handling multiple closely spaced GEO objects. Li et al.[6] constructed association matrices from initial association results and employed the bond energy algorithm for clustering. Additionally, a splitting algorithm was proposed to reduce computational complexity and mitigate erroneous associations. However, it may cause association errors for space objects with highly similar orbits.

In summary, tracklet clustering is related to tracklet association, as clustering relies heavily on the association results. However, traditional clustering methods tend to be overly dependent on the association results. With the increasing deployment of satellite constellations and formation-flying missions, the accuracy of tracklet association is expected to deteriorate. Moreover, the increasing number of space objects will intensify the computational burden, underscoring the need for more robust and scalable clustering methodologies.

In response to these needs, this work proposes a probabilistic and optimization-based approach for tracklet clustering and initial orbit determination. The method aims to efficiently group tracklets under uncertainty, leveraging association probabilities derived from angular-rate consistency. This enables both accurate object reconstruction and scalable catalog maintenance for future SSA systems.

The primary contributions of this paper are:

- A novel Probabilistic Tracklet Clustering (PTC) framework that naturally handles association uncertainty without arbitrary thresholds.
- A scalable optimization-based method to compute the most probable clustering configuration.
- A lightweight, probabilistic IOD strategy that effectively combines multiple partial orbit estimates.

The remainder of this paper is organized as follows. Sec. 2 introduces the optical measurement model and formally defines the tracklet association problem. Sec. 3 details the proposed PTC algorithm, including the association probability model, clustering formulation, and optimization procedure. Sec. 4 describes the simulation setup and evaluates the proposed framework through a set of numerical experiments. Finally, Sec. 5 summarizes the main findings and outlines future research directions.

II. Background

This section outlines the optical measurement framework utilized to acquire angular motion attributes of RSOs. What's more, the concrete problem of tracklet association and clustering is stated.

II.i Measurement Model

The measurement model assumes that the observers directly acquire target angular information in the observer's topocentric spherical coordinate frame. Angular velocity is then derived from a series of measurements.

Suppose observers acquire sequential measurements $\mathcal{Z}$ of topocentric right ascension (α) and declination (δ) at discrete time t_i ($i = 1, 2, ..., N$).

$$\mathcal{Z} = \left\{ \left(\alpha(t_i), \delta(t_i)\right) \mid t_1 \leq t_i \leq t_N \right\}. \qquad [1]$$

The measurements α and δ can be calculated in the Earth-Centered Inertial (ECI) coordinate frame through the geometric relationship between the target and the observer.[13]

$$\begin{aligned} \alpha &= \arctan\left(\frac{r_y - r_{y,O}}{r_x - r_{x,O}}\right), \\ \delta &= \arcsin\left(\frac{r_z - r_{z,O}}{\|\boldsymbol{r} - \boldsymbol{r}_O\|}\right), \end{aligned} \qquad [2]$$

where $\boldsymbol{r} = (r_x, r_y, r_z)^T$ and $\boldsymbol{r}_O = (r_{x,O}, r_{y,O}, r_{z,O})^T$ denote the location vectors of the target and observer, respectively, in the ECI coordinate frames.

Given a set of angle measurements $\mathcal{Z}$, the angular rate $\dot{\alpha}, \dot{\delta}$ can be estimated using the least squares algorithm developed by Maruskin.[17] The uncertainty of angular velocity can also be estimated:[17]

$$\dot{\sigma} \sim \frac{2\sqrt{3}\sigma}{T\sqrt{N}}, \qquad [3]$$

where σ is the standard deviation of the angle, T is the observation time, N is the number of observations. Then, combine the angle and angular velocity to form the optical attributable vector, i.e., tracklet:

$$A = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) \in [-\pi, \pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times \mathbb{R}^2. \qquad [4]$$

II.ii Problem Statement

In optical Space Surveillance and Tracking (SST) systems, maintaining an accurate space object catalog involves two core tasks: updating the orbits of known Resident Space Objects (RSOs) and determining the orbits of newly detected ones. The surveillance systems generate thousands of new tracklets daily. Initially, these tracklets are associated with cataloged RSOs by comparing them with simulated observations through orbit propagation models, which is called Tracklet-To-Orbit Association (T2OA). Specifically, the predicted positions and observation attributes of cataloged RSOs are computed at the epochs of the tracklets using orbital propagators. The association process involves evaluating the consistency between the observed tracklets and the simulated observation using statistical metrics.[18,19]

If a tracklet is successfully associated with a cataloged RSO, it is treated as an additional observation that can be assimilated into an orbit determination (OD) process. This refined OD process leverages batch least-squares estimation or sequential filters,[20] such as the Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF), to update the state vector and covariance of the associated object, thereby enhancing orbital accuracy. The updated orbital state is then re-integrated forward to maintain the ephemeris of the RSO in the catalog for future observation planning and correlation. Conversely, if a tracklet cannot be confidently associated with any cataloged object, it is classified as an Uncorrelated Tracklet (UCT) and marked for further analysis.

The further analysis of UCTs typically proceeds through three main stages: tracklet association, tracklet clustering, and maneuver detection. Tracklet association different from T2OA, also known as Tracklet-To-Tracklet Association (T2TA), seeks to identify whether two short arcs originate from the same RSO. If so, they are stored as a matched pair.

Tracklet clustering, which is the focus of this study, builds upon these pairwise associations to group multiple tracklets belonging to the same object. This stage aims to minimize errors from false associations while ensuring scalability and computational efficiency. Once clustering is completed, precise orbit determination and maneuver detection can be carried out. These procedures facilitate the cataloging of new RSOs or the identification of maneuvers from existing ones. A newly confirmed object results in the creation of a new catalog entry, whereas a detected maneuver prompts an update to the corresponding RSO's orbital parameters. Together, tracklet association and clustering constitute the essential preparatory steps for any subsequent maneuver analysis.

III. Method

This section outlines the methodology for tracklet association, probabilistic clustering, and initial orbit determination in the context of space object tracking. The proposed framework is based on the BVP-CAR-Opt association method, which estimates orbits from pairs of tracklets by solving a constrained optimization problem. The resulting association loss function is then transformed into a probabilistic similarity measure to guide the clustering process. A greedy algorithm is employed to construct cluster partitions that maximize the overall association likelihood while satisfying structural constraints. Finally, the orbit states within each cluster are fused using a probabilistic weighting scheme to generate initial orbit estimates. The following subsections describe each component of the proposed method in detail.

III.i Tracklet Association

The method for tracklet association utilized here is BVP-CAR-Opt (Optimization), which improved the BVP-CAR association method and can estimate orbits of two tracklets.

Given two tracklets and the assumed ranges of tracklets, the targets' positions $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$ at the observation times can be calculated as:

$$\boldsymbol{r}^{(i)} = \boldsymbol{r}_O^{(i)} + \rho^{(i)}\boldsymbol{u}^{(i)}, \quad i = 1, 2, \tag{5}$$

where $\boldsymbol{r}_O^{(i)}$ is the observer's position, $\rho^{(i)}$ is the range, and $\boldsymbol{u}^{(i)}$ is the line-of-sight unit vector derived from angular measurements.

The positions can be utilized to solve the Lambert problem, yielding velocity vectors $\boldsymbol{v}^{(1)}$, $\boldsymbol{v}^{(2)}$. Predicted angular rates $\hat{\dot{\alpha}}, \hat{\dot{\delta}}$ are computed from the orbit states $\boldsymbol{X}^{(i)} = [\boldsymbol{r}^{(i)T}, \boldsymbol{v}^{(i)T}]^T$, and the association is evaluated by minimizing the loss function:

$$f(\rho^{(1)}, \rho^{(2)}) = (\dot{\boldsymbol{y}} - \hat{\dot{\boldsymbol{y}}})^T (C_{\dot{\boldsymbol{y}}} + C_{\hat{\dot{\boldsymbol{y}}}})^{-1} (\dot{\boldsymbol{y}} - \hat{\dot{\boldsymbol{y}}}), \tag{6}$$

where $\dot{\boldsymbol{y}}$ and $\hat{\dot{\boldsymbol{y}}}$ are the measured and predicted angular rates, and $C_{\dot{\boldsymbol{y}}}$ and $C_{\hat{\dot{\boldsymbol{y}}}}$ their respective covariances.

The optimization is subject to several constraints ensuring orbital feasibility, including:[10]

- inclination bounds,
- Lambert solution constraints,
- zero-energy time-of-flight feasibility,
- vacant focus geometry conditions, and
- monotonicity of time-of-flight.

Tracklets are considered associated if a feasible solution exists.

III.ii Probabilistic Tracklet Clustering

The loss function value of BVP-CAR-Opt can reflect the association probability between tracklet pairs, to some extent. In general, a higher loss value indicates a lower probability of association, whereas a lower loss value suggests a higher probability. This characteristic allows the loss function to serve as an implicit measure of confidence in tracklet association. Specifically, for a given tracklet pair (A_i, A_j) with an associated loss f_{ij}, the association probability is defined as:

$$P_{ij} = \exp(-\lambda f_{ij}), \quad [7]$$

where $\lambda > 0$ is a scaling parameter controlling the sharpness of the probability distribution. This exponential transformation ensures that higher-quality associations (lower loss values) are assigned higher probabilities.

Let $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ denote a set of n tracklets (optical attributables). A valid clustering of $\mathcal{A}$ corresponds to a partition $\{C_1, C_2, \ldots, C_k\}$ into k non-empty, disjoint subsets, subject to the following structural constraints:[21]

1. **Completeness:** The union of all clusters must cover the entire set:

$$\bigcup_{i=1}^{k} C_i = \mathcal{A}.$$

2. **Disjointness:** Any two clusters must be mutually exclusive:

$$C_i \cap C_j = \emptyset, \quad \forall i \neq j.$$

3. **Non-emptiness:** Each cluster must contain at least one element:

$$|C_i| \geq 1, \quad \forall i = 1, \ldots, k.$$

The total number of such partitions of an n-element set is given by the Bell number B_n,[22] which grows super-exponentially with n:

$$B_n = \sum_{k=0}^{n} \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n. \quad [8]$$

What's more, each partition of $\mathcal{A}$ can be equivalently represented by an $n \times n$ binary association matrix $\mathbf{M} \in \{0,1\}^{n \times n}$, where:

$$M_{ij} = \begin{cases} 1, & \text{if } A_i \text{ and } A_j \text{ belong to the same cluster,} \\ 0, & \text{otherwise.} \end{cases}$$

The matrix $\mathbf{M}$ must satisfy the properties of an equivalence relation:

- **Reflexivity:** $M_{ii} = 1$ for all i,
- **Symmetry:** $M_{ij} = M_{ji}$ for all i, j,
- **Transitivity:** If $M_{ij} = 1$ and $M_{jk} = 1$, then $M_{ik} = 1$.

Thus, each feasible partition of the tracklet set corresponds uniquely to a symmetric, reflexive, and transitive binary matrix $\mathbf{M}$, which defines the clustering result for association analysis.

While Bell enumeration captures all possible clustering outcomes, its computational complexity is intractable for even moderate n due to its super-exponential growth. For example, $B_{10} = 115\,975$ and $B_{15} \approx 1.38 \times 10^9$, rendering exhaustive evaluation infeasible in practical applications.

To circumvent the combinatorial explosion, the clustering task is reformulated as a constrained optimization problem. Let $\mathcal{M} = \{\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_K\}$ denote a (partial or sampled) set of candidate clustering matrices. Define the probability of clustering configuration $\mathbf{M}_k$ as:

$$\mathbb{P}(\mathbf{M}_k) = \prod_{i<j} P_{ij}^{M_{ij}^{(k)}} (1 - P_{ij})^{1 - M_{ij}^{(k)}}, \quad [9]$$

To identify the most probable clustering configuration, we solve the following optimization problem:

$$\begin{aligned} \max_{\mathbf{M} \in \mathcal{C}} \quad & \sum_{i<j} \left[M_{ij} \log P_{ij} + (1 - M_{ij}) \log(1 - P_{ij}) \right], \\ \text{s.t.} \quad & \mathbf{M} \text{ satisfies symmetry and transitivity,} \\ & M_{ij} \in \{0,1\}, \quad \forall i, j. \end{aligned} \quad [10]$$

Here, $\mathcal{C}$ denotes the set of feasible clusters that satisfy the above constraints.

This probabilistic clustering framework allows for a principled handling of uncertainties in tracklet association without relying on user-defined thresholds.

III.iii Optimization Method

To efficiently solve the constrained clustering optimization problem introduced in Section 3.2, a greedy optimization algorithm is adopted in this work. The core idea is to iteratively merge tracklet clusters in a manner that maximally increases the overall clustering likelihood while preserving the structural constraints of valid clustering matrices. Specifically, the algorithm proceeds as follows:

1. **Initialization:** Each tracklet is assigned to its singleton cluster, corresponding to an identity matrix $\mathbf{M}_0 = \mathbf{I}_n$.
2. **Pairwise Gain Evaluation:** For each pair of current clusters (C_a, C_b), compute the total score

gain Δ_{ab} that would result from merging them. This gain is computed using the objective function:

$$\Delta_{ab} = \sum_{i \in C_a} \sum_{j \in C_b} \log\left(\frac{P_{ij}}{1 - P_{ij}}\right).$$

3. **Greedy Merge:** Select the cluster pair (C_a, C_b) with the highest positive Δ_{ab} and merge them. Update the cluster labels and the binary matrix $\mathbf{M}$ accordingly.

4. **Termination:** Repeat steps 2–3 until no further positive gain can be obtained or the iteration number has exceeded the max iteration number $K_{\max}$.

This strategy ensures that each merge operation strictly improves the log-likelihood of the current clustering configuration. The transitivity constraint is naturally preserved by merging entire clusters instead of individual elements. Symmetry and reflexivity are also maintained by the construction of the matrix $\mathbf{M}$ during the merge process.

Computational Complexity: The greedy algorithm has a worst-case time complexity of $\mathcal{O}(n^3)$, arising from evaluating all $\mathcal{O}(n^2)$ cluster pairs in each of up to $\mathcal{O}(n)$ iterations. However, in practice, the number of iterations is typically much smaller due to the diminishing number of available merges and early termination conditions.

Overall, the greedy optimization approach strikes a balance between computational efficiency and clustering quality, making it suitable for real-world tracklet association scenarios in SSA.

III.iv Initial Orbit Determination

Once the tracklet clustering results are obtained, along with the pairwise orbit determinations and their associated probabilities, a fused orbit state for each cluster can be estimated. The fusion is performed via a probabilistic weighted average, without requiring the covariance of the estimated orbits.

Let $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ denote the set of k clusters, where each cluster C_l contains a subset of tracklet indices. For each cluster $C_l = \{i_1, i_2, \ldots, i_m\}$, we define the fused orbit state vector $\boldsymbol{X}^{(l)} \in \mathbb{R}^6$ as:

$$\boldsymbol{X}^{(l)} = \frac{\sum_{p<q} P_{i_p i_q} \boldsymbol{X}_{i_p i_q}}{\sum_{p<q} P_{i_p i_q}}, \qquad [11]$$

where:

- $\boldsymbol{X}_{i_p i_q} \in \mathbb{R}^6$ is the estimated orbit state obtained from tracklet pair (i_p, i_q);
- $P_{i_p i_q} \in [0, 1]$ is the corresponding association probability;
- The summation is taken over all unordered pairs within the cluster: $1 \le p < q \le m$.

If a particular pair (i_p, i_q) does not yield a valid orbit estimate (i.e., $\boldsymbol{X}_{i_p, i_q}$ is undefined), it is excluded from the computation. The final fused state $\boldsymbol{X}^{(l)}$ represents the consensus orbit of the objects within cluster C_l, based on probabilistic confidence in each pairwise association.

This approach provides a lightweight, non-iterative means of aggregating multiple pairwise orbit estimates into a single state vector, effectively leveraging the probabilistic output of the BVP-CAR-Opt association method.

IV. Experimental Results

In this section, a simulated case study is investigated to validate the effectiveness of the proposed algorithm. Sec. 4.1 details the scenario settings. In Sec. 4.2, five evaluation metrics of clustering are presented. Sec. 4.3 discusses the results of the algorithm.

IV.i Simulation Design

The simulation scenario, as shown in Fig. 1, wherein a LEO satellites sequentially observe 30 other GEO satellites by sensor tasking. The initial orbital conditions for both the target and the observer satellite are generated using Two-Line Element (TLE) data from the space object catalog maintained by the 18th Space Defense Squadron*, while the actual states of satellites during the simulation are computed using an orbital propagator in the ECI coordinate frame. The initial orbital elements of the observer satellite at the starting epoch are presented in Table 1. Furthermore, it is assumed that the observer satellite's orbital states can be accurately obtained using GNSS measurements. The simulation begins at 2024-01-01T21:00:00Z (UTC) and lasts for 6 hours, until 2024-01-02T03:00:00Z (UTC), generating 97 tracklets with attributable vectors (including right ascension, declination, right ascension rate, and declination rate). The measurements are randomly generated following a Gaussian distribution, with the true value as the mean[23] and a standard deviation calculated by:[17]

*Data available online at http://www.space-track.org/[retrieved April 1, 2024]

$$\begin{aligned}\sigma_\alpha &= \sigma_\delta = 1'',\\ \dot{\sigma}_\alpha &= \dot{\sigma}_\delta = 0.0036''/\text{s}.\end{aligned} \qquad [12]$$

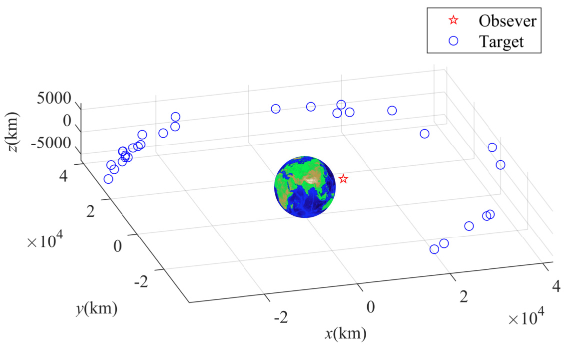

Fig. 1: The scenario at the initial moment

GEO objects are positioned far from Earth so that they are less affected by perturbations over a short period. Given the short time intervals between the tracklets of GEO objects, perturbation effects are negligible. This allows for the use of a simplified Kepler two-body dynamic model, which enhances computational efficiency without significantly affecting association accuracy.

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r}. \qquad [13]$$

All simulations were conducted on a computer with an Intel i7-12700K 3.60 GHz CPU.

IV.ii Evaluation Metrics

To assess the effectiveness of the tracklet clustering algorithm, we adopt several widely used external evaluation metrics. These metrics compare the predicted cluster labels with the ground truth identities and quantify the clustering quality from different perspectives, such as accuracy, consistency, and mutual information. The metrics used in this study are as follows:[24]

1. **Purity**: Purity evaluates the extent to which each predicted cluster contains samples from a single ground truth class. It is defined as:

$$\text{Purity} = \frac{1}{M}\sum_{k=1}^{K}\max_j |C_k \cap L_j|, \qquad [14]$$

where M is the total number of samples, C_k is the set of samples in predicted cluster k, and L_j is the set of samples in ground truth class j.

2. **Rand Index (RI)**: The Rand Index measures the agreement between two partitions by considering all sample pairs and counting how many pairs are assigned consistently in both the predicted and ground truth partitions. It is defined as:

$$\text{RI} = \frac{TP + TN}{TP + FP + FN + TN}, \qquad [15]$$

where TP, TN, FP, and FN denote the number of true positive, true negative, false positive, and false negative pairs, respectively.

3. **F1-score**: The F1-score is the harmonic mean of precision and recall, balancing both metrics in one measure. It is calculated as:

$$F_1 = \frac{2TP}{2TP + FP + FN}. \qquad [16]$$

4. **Normalized Mutual Information (NMI)**: NMI quantifies the mutual dependence between the clustering result and the ground truth labels using concepts from information theory. It is defined as:

$$\text{NMI} = \frac{2 \cdot I(U;V)}{H(U) + H(V)}, \qquad [17]$$

where $I(U;V)$ is the mutual information between predicted clustering U and ground truth V, and $H(U)$, $H(V)$ are the entropies of their respective label distributions.

5. **Fowlkes–Mallows Index (FM)**: FM Index evaluates clustering quality based on the geometric mean of precision and recall at the pairwise level. It is given by:

$$\text{FM} = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}. \qquad [18]$$

IV.iii Results

To evaluate the effectiveness of the proposed method, we compare it with traditional clustering techniques, including the classical MCL algorithm and its weighted variant (Weighted MCL). The performance is assessed both visually using confusion matrices (after label alignment via the Hungarian algorithm) and quantitatively through several standard clustering metrics. Fig. 2, Fig. 3, and Fig. 4 present the confusion matrices of the clustering results for all

Table 1: The initial orbital elements of the observer satellites

Satellite	a (ER)	e	i (deg)	Ω (deg)	ω (deg)	M (deg)
Observer	1.2741	0.1461	32.8836	12.7823	146.9093	-136.8638

methods after label alignment. The PTC achieves a more diagonal-dominant structure, indicating better agreement between the predicted cluster labels and the ground truth.

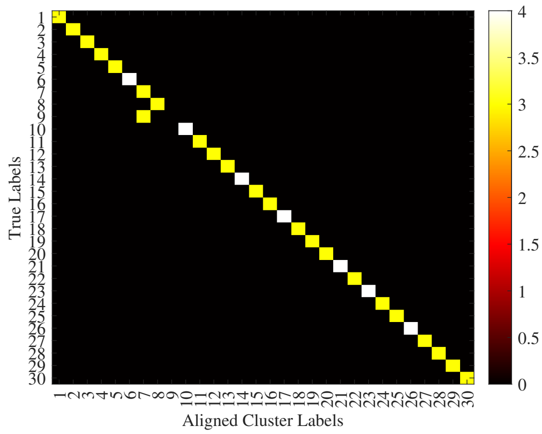

Fig. 2: Aligned confusion matrix for MCL

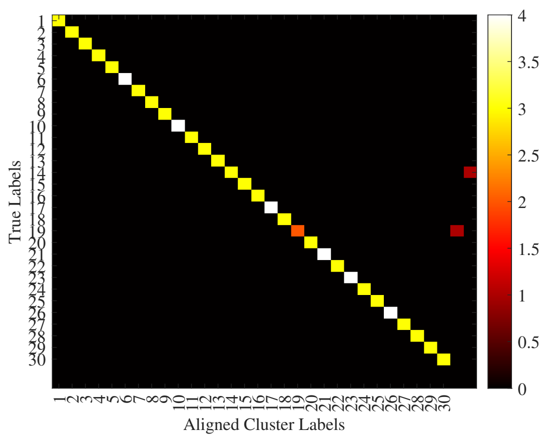

Fig. 3: Aligned confusion matrix for Weighted MCL

To further evaluate the clustering performance, we compare the results of MCL, Weighted MCL, and the PTC in terms of Purity, RI, F1, NMI, and FM. As shown in Table 2, the PTC achieves the highest scores across all metrics, consistently outperforming the baseline methods and demonstrating its effectiveness. As observed, the improvements are especially significant in NMI and F1-score, which are sensitive to both cluster assignment accuracy and information preservation. This suggests that the PTC not only matches ground truth clusters better but also provides more coherent and separable groupings.

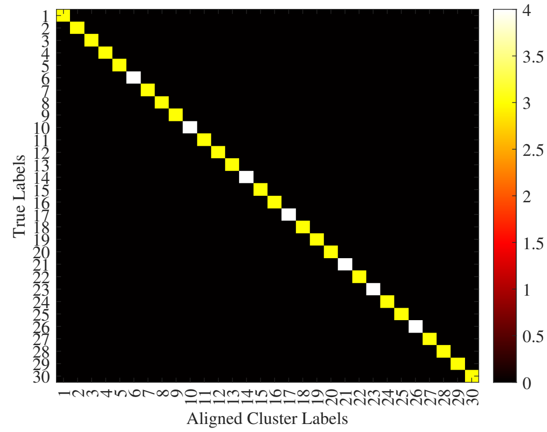

Fig. 4: Aligned confusion matrix for PTC

To evaluate the accuracy of the estimated orbits, the fused orbit states $\boldsymbol{X}^{(l)}$ are compared against the ground truth orbital states of the corresponding GEO targets used in the simulation. Table 3 summarizes the root mean square (RMS) errors in position and velocity across all 30 clusters. For comparison, orbit estimates obtained using naive averaging (i.e., equal-weighted fusion) are also computed, and their corresponding performance is reported. As shown in Table 3, the proposed IOD strategy reduces both position and velocity errors compared to the equal-weighted baseline. This demonstrates that incorporating association probabilities in the fusion process enhances the accuracy and consistency of the resulting orbit estimates.

V. Conclusions

This paper presents a probabilistic framework for tracklet clustering and IOD, addressing the challenges posed by short-arc optical observations and the

Table 2: Quantitative comparison of clustering performance

Method	Purity	Rand Index (RI)	F1-score	NMI	FM
MCL	0.9691	0.9981	0.9610	0.9618	0.9937
Weighted MCL	1	0.9989	0.9770	0.9772	0.9937
PTC	**1**	**1**	**1**	**1**	**1**

Table 3: RMS errors of initial orbit determination (IOD) in position and velocity domains

Method	Position Error (km)	Velocity Error (m/s)
Equal-weighted averaging	11.04	1.25
Proposed probabilistic fusion	**0.35**	**0.03**

increasing density of RSOs. By integrating the BVP-CAR-Opt with a probabilistic association model and formulating clustering as a constrained optimization problem, the PTC effectively balances accuracy, robustness, and computational efficiency. The experimental results demonstrate that the proposed algorithm outperforms traditional methods, such as MCL and Weighted MCL, across multiple evaluation metrics. Moreover, the lightweight IOD strategy, which fuses multiple partial orbit estimates based on association probabilities, further enhances catalog initialization. This work contributes a scalable and principled approach for improving space object cataloging. In the future, the probabilistic clustering framework could be extended to incorporate temporal constraints or learning-based priors to further reduce false associations and improve clustering reliability in dense and complex orbital environments.

References

[1] Jingrui Zhang, Yifan Cai, Chenbao Xue, Zhirun Xue, and Han Cai. Leo mega constellations: review of development, impact, surveillance, and governance. *Space: Science & Technology*, 2022.

[2] Jin Choi, Jung Hyun Jo, Hong-Suh Yim, Eun-Jung Choi, Sungki Cho, and Jang-Hyun Park. Optical tracking data validation and orbit estimation for sparse observations of satellites by the owl-net. *Sensors*, 18(6):1868, 2018.

[3] Zhe Zhang, Gaopeng Zhang, Jianzhong Cao, Cheng Li, Weining Chen, Xin Ning, and Zheng Wang. Overview on space-based optical orbit determination method employed for space situational awareness: From theory to application. In *Photonics*, volume 11, page 610. MDPI, 2024.

[4] Sajjad Kazemi, Nasser L Azad, K Andrea Scott, Haroon B Oqab, and George B Dietrich. Orbit determination for space situational awareness: A survey. *Acta Astronautica*, 2024.

[5] Xingyu Zhou, Shuo Wang, and Tong Qin. Multi-spacecraft tracking and data association based on uncertainty propagation. *Applied Sciences*, 12(15), 2022.

[6] Jiasheng Li, Zhen Yang, and Yazhong Luo. Too-short-arc association and clustering method for space-based optical observations. *Journal of Guidance, Control, and Dynamics*, 48(1):84–97, 2025.

[7] JA Siminski, Oliver Montenbruck, Hauke Fiedler, and Thomas Schildknecht. Short-arc tracklet association for geostationary objects. *Advances in space research*, 53(8):1184–1194, 2014.

[8] Han Cai, Yang Yang, Steve Gehly, Suqin Wu, and Kefei Zhang. Improved tracklet association for space objects using short-arc optical measurements. *Acta Astronautica*, 151:836–847, 2018.

[9] JA Gaebler, P Axelrad, and PW Schumacher Jr. Boundaries on range-range constrained admissible regions for optical space surveillance. In *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, page 12, 2017.

[10] Paul W Schumacher, John A Gaebler, Christopher WT Roscoe, Matthew P Wilkins, and Penina Axelrad. Parallel initial orbit determination using angles-only observation pairs. *Ce-*

lestial Mechanics and Dynamical Astronomy, 130(9):60, 2018.

[11] Yang Yang, Zhenwei Li, Han Cai, and H Zhang. Single-/multiple-revolution lambert's problem constraints for optical track-to-track association. *Advances in the Astronautical Sciences AAS/AIAA Astrodynamics*, 177, 2021.

[12] Laura Pirovano, Gennaro Principe, and Roberto Armellin. Data association and uncertainty pruning for tracks determined on short arcs. *Celestial Mechanics and Dynamical Astronomy*, 132:1–23, 2020.

[13] Xuejian Mao, Zihan Zhou, and Pei Chen. Optical tracklet-to-tracklet correlation with control-distance metrics. *Journal of Guidance, Control, and Dynamics*, 48(3):646–655, 2025.

[14] S. van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, Utrecht University, Utrecht, The Netherlands, 2000.

[15] Carlos Yanez, Juan Carlos Dolado Pérez, Pascal Richard, Ivan Llamas, and Laurent Lapasset. Optical measurements association using optimized boundary value initial orbit determination coupled with Markov clustering algorithm. In *7th European Conference on Space Debris*, Darmstadt, Germany, Apr 2017. ESA.

[16] T. Zhu, X. Wang, J. Zhang, S. Yu, and I. Molotov. Mean-shift clustering approach to the tracklets association with angular measurements of resident space objects. *Astronomy and Computing*, 40:100588, 2022.

[17] Jared M Maruskin, Daniel J Scheeres, and Kyle T Alfriend. Correlation of optical observations of objects in earth orbit. *Journal of Guidance, Control, and Dynamics*, 32(1):194–209, 2009.

[18] Michiel Zittersteijn, Alessandro Vananti, Thomas Schildknecht, JC Dolado Perez, and V Martinot. Associating optical measurements of meo and geo objects using population-based meta-heuristic methods. *Advances in space research*, 58(9):1778–1792, 2016.

[19] Alessandro Vananti, Moritz Meyer zu Westram, and Thomas Schildknecht. Metrics on space of closed orbits for near-earth objects identification. *Celestial Mechanics and Dynamical Astronomy*, 135(5):51, 2023.

[20] J.R. Raol and N.K. Sinha. On the orbit determination problem. *IEEE Transactions on Aerospace and Electronic Systems*, AES-21(3):274–291, 1985.

[21] Jiawei Han, Jian Pei, and Hanghang Tong. *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[22] Nobuhiro Asai, Izumi Kubo, and Hui-Hsiung Kuo. Bell numbers, log-concavity, and log-convexity. *Acta Applicandae Mathematica*, 63(1-3):79–87, 2000.

[23] Sebastian Samuele Rizzuto, Riccardo Cipollone, Andrea De Vittori, Pierluigi Di Lizia, and Mauro Massari. Object detection on space-based optical images leveraging machine learning techniques. *Neural Computing and Applications*, pages 1–25, 2025.

[24] Sami Abdulla Mohsen Saleh, A Halim Kadarman, Shahrel Azmin Suandi, Sanaa AA Ghaleb, Waheed AHM Ghanem, Solehuddin Shuib, and Qusay Shihab Hamad. A tracklet-before-clustering initialization strategy based on hierarchical klt tracklet association for coherent motion filtering enhancement. *Mathematics*, 11(5):1075, 2023.