

Using Apparent Motion Parameters as Features for Machine Learning-based Tracklet Association: An Analysis of Feature Discriminative Power

T. N. Sannikova

Abstract. The association of short observational arcs (tracklets) becomes critically challenging in spacecraft fragmentation events, where multiple fragments move along closely spaced trajectories. Traditional orbit-determination-based methods generate numerous false associations in such conditions. This study explores geometric features derived directly from positional measurements $\alpha(t)$, $\delta(t)$ and apparent motion parameters (angular rate, position angle, small-circle parameters). We compute features for each pair, analyze their separability, and identify the most informative ones. The selected set includes six approximation residuals and three geometric criteria: the alignment angle α_{align} , the separation ratio R_{sep} , and the position angle difference $\Delta\psi$. Other parameters were discarded due to overlapping distributions or noise sensitivity. Three classifiers (Decision Tree, Random Forest, LightGBM) trained on these features demonstrate high performance. This approach provides reliable association without preliminary orbit determination, making it suitable for real-time applications.

Introduction

The association of short observational arcs (tracklets) is a fundamental problem in space surveillance, particularly critical in spacecraft fragmentation events where multiple debris fragments move along closely spaced trajectories. Classical methods based on preliminary orbit determination often generate numerous false associations due to orbital solution degeneracy, and require high computational cost. As an alternative, geometric characteristics of apparent motion offer a promising source of information. Modern CCD observations provide dense coordinate series $\alpha(t)$, $\delta(t)$, enabling the direct computation of apparent motion parameters (AMPs) without orbit propagation. In this paper, we analyze the separability of

various features derived from positional measurements for pairwise tracklet classification, select a compact nine-feature set, and validate its performance using machine learning models.

1. Features for a Pair of Tracklets

The dataset comprises 840 tracklets obtained during 2022–2023 at the Crimean Astrophysical Observatory. 363 tracklets belong to 168 distinct objects observed over one night (2–7 tracklets per object); the remaining tracklets from the same nights provide diverse false associations. Pairs were formed with two constraints: no temporal overlap, and a time difference not exceeding 10 hours, yielding 5315 pairs (238 true associations, 5077 false associations).

Following [1], for each tracklet containing n points $(t_j, \alpha_j, \delta_j)$, we compute AMPs for the mean epoch $t_0 = (t_1 + t_n)/2$: position vector $\mathbf{D}$, angular rate μ , position angle ψ , geodesic curvature κ , normal $\mathbf{P}$ and distance p of the osculating small circle, and tangent vector $\mathbf{T}$.

For each pair of tracklets A and B , the following features are computed:

1. **Joint Approximation Residuals:** Six RMS residuals $\sigma_{\text{line}}, \sigma_\alpha, \sigma_\delta, \sigma_{D1}, \sigma_{D2}, \sigma_{D3}$ from second-degree polynomial fits to $\delta(\alpha)$, $\alpha(t)$, $\delta(t)$, and the direction cosines $D_1 = \cos \delta \cos \alpha$, $D_2 = \cos \delta \sin \alpha$, $D_3 = \sin \delta$.
2. **Alignment Angle** between the displacement vector and the mean tangent:

$$\alpha_{\text{align}} = \frac{180}{\pi} \arccos \left(\frac{|(\mathbf{D}_B - \mathbf{D}_A) \cdot \mathbf{T}_{\text{mean}}|}{|\mathbf{D}_B - \mathbf{D}_A|} \right) \quad [\text{deg}], \quad \mathbf{T}_{\text{mean}} = \frac{\mathbf{T}_A + \mathbf{T}_B}{|\mathbf{T}_A + \mathbf{T}_B|}.$$

3. **Separation Ratio** of measured angular separation to expected separation:

$$R_{\text{sep}} = \frac{\rho_{AB}}{\frac{1}{2}(\mu_A + \mu_B) \Delta t_{AB}},$$

where Δt_{AB} is the time interval between mean epochs and ρ_{AB} is the angular separation between mean positions.

4. **Position Angle Difference** (reduced to $[0^\circ, 180^\circ]$):

$$\Delta \psi = \frac{180}{\pi} \min(|\psi_A - \psi_B|, 2\pi - |\psi_A - \psi_B|) \quad [\text{deg}].$$

5. **Consistency of planes of small circles:**

$$\theta_P = \arccos(|\mathbf{P}_A \cdot \mathbf{P}_B|) \quad [\text{deg}], \quad \Delta p = |p_A - p_B|.$$

6. **Difference in angular velocities** $\Delta \mu = |\mu_A - \mu_B| \quad [\text{angular sec/s}]$.

7. **Difference in geodesic curvatures** $\Delta \kappa = |\kappa_A - \kappa_B|$.

In addition, the time interval Δt_{AB} and the angular separation ρ_{AB} were evaluated as auxiliary features. Their individual distributions inherently overlap strongly, which is expected for representative observational data; nevertheless, they were tested in combination with the primary features to explore whether they could provide complementary discriminative information.

2. Separability Analysis and Feature Selection

Analysis of the training set reveals that α_{align} exhibits the highest separability: 76% of true pairs have α_{align} below 0.1° (only 0.4% for false), and all true pairs satisfy $\alpha_{\text{align}} \leq 0.73^\circ$. R_{sep} also performs strongly, with 91% of true pairs lying within (0.90, 1.01), compared to only 5% of false pairs. The position angle difference $\Delta\psi$ typically localizes near 0° for true pairs. However, due to the topocentric effect (loops in the apparent trajectory), larger values are possible — a case of $\Delta\psi \approx 95^\circ$ was recorded in the training sample. Nevertheless, 71% of true pairs and only 24% of false pairs have $\Delta\psi < 5^\circ$.

The joint approximation residuals provide strong discriminative power. For a threshold of $30''$, σ_{line} is below the threshold for 96.2% of true pairs but only for 11.5% of false pairs. The remaining five residuals exhibit a similar pattern: the fraction of true pairs with residual $\leq 30''$ ranges from 71% to 91%, while for false pairs it never exceeds 12%. This confirms that the smoothness of the combined trajectory is a reliable indicator of a true association.

However, several investigated features were excluded from the final set: the angle between small-circle normals θ_P and the distance difference Δp proved too sensitive to measurement noise on short arcs, with strongly overlapping distributions. The differences in angular velocity $\Delta\mu$ and geodesic curvature $\Delta\kappa$ also showed significant overlap between classes. Furthermore, the inclusion of the time interval Δt_{AB} and angular separation ρ_{AB} into the feature set did not yield any improvement in the classification metrics; consequently, they were excluded.

Thus, the final feature set comprises six approximation residuals (σ_{line} , σ_α , σ_δ , σ_{D1} , σ_{D2} , σ_{D3}) and three geometric criteria (α_{align} , R_{sep} , $\Delta\psi$).

3. Machine Learning Models and Feature Importance

Three classifiers were trained on the selected features: Decision Tree (DT), Random Forest (RF), and LightGBM gradient boosting. Class imbalance was compensated by weighting true pairs inversely to their proportion. The full dataset was split into training and test subsets in a 75:25 ratio, and hyperparameters were tuned via cross-validation to maximize the area under the precision-recall curve (AUC-PR). Each trained classifier outputs the probability p_{AB} that a pair of tracklets belongs to the same physical object; a pair is considered a true association if $p_{AB} \geq 0.5$.

Feature importance analysis shows that α_{align} is dominant in all models. In DT, it accounts for 92.4% of total importance, with R_{sep} as the second most important feature at 7.3%. In RF, α_{align} decreases to 42.2%, while approximation residuals σ_{line} (19.6%), σ_{D2} (16.2%), and σ_{D3} (8.5%) become significant. In LightGBM, the contributions are more evenly distributed: α_{align} (21.9%), R_{sep} (15.7%), σ_{D2} (13.1%), and σ_α (11.8%). The use of multiple parameters allows ensemble methods (Random Forest and LightGBM) to achieve a better precision-recall balance compared to the decision tree, which relies almost entirely on α_{align} .

The high AUC-PR values (≥ 0.98) obtained for all three classifiers confirm strong discriminative power across all thresholds p_{AB} . At the 0.5 threshold, Random Forest missed 2 true pairs, while the other models achieved perfect recall (1.0). The difference in false positives is relatively small: Random Forest performed best with only 6 incorrectly linked pairs and a precision of 0.906; Decision Tree made 8 errors, and LightGBM produced 9 false positives, reducing its precision to 0.870.

Conclusion

The method was tested on 20 tracklets obtained during an 80-minute observation session of debris fragments from the ATLAS 5 CENTAUR R/B upper stage breakup in September 2024. The target set contains 7 objects with 2 to 5 tracklets each, plus 2 isolated tracklets, yielding 183 possible pairs (18 true, 165 false). On the target dataset, Decision Tree yielded the fewest false positives (19) while achieving 100% recall; Random Forest and LightGBM also achieved 100% recall with 20 and 37 false positives, respectively.

After applying maximum weight matching in a bipartite graph to form consistent chains of tracklets, all three classifiers yield the correct partition on the target data. On the full dataset (168 chains (each with 2 to 7 tracklets) and 477 single tracklets), Random Forest and LightGBM correctly reconstructed 166 complete chains (98.8%), while Decision Tree reconstructed 165 (98.2%). Random Forest produced the fewest false merges (4) compared to 6 (LightGBM) and 11 (Decision Tree), but fragmented 2 chains; the other two models did not fragment any chains.

This study shows that approximation residuals that characterize trajectory smoothness and geometric criteria based on apparent motion parameters provide effective discrimination for tracklet association. The proposed approach requires no preliminary orbit determination, making it computationally efficient and suitable for real-time applications in space surveillance and fragmentation monitoring.

The author thanks Vasily V. Rumyantsev, senior researcher at CrAO RAS, for providing the observational data.

The study was carried out within the state assignment of the Federal State Budget Scientific Institution "Crimean Astrophysical Observatory of RAS".

References

- [1] Bykov O.P., Kholshevnikov K.V., *Direct methods for determining the orbits of celestial bodies: a tutorial*, Spb: Izd-vo S.-Peterb. un-ta, 2013 [in Russian].

T. N. Sannikova

Laboratory of small solar system bodies of the Solar Physics and Solar System Department

Crimean Astrophysical Observatory of the Russian Academy of Sciences

Nauchny, Crimea, Russia

e-mail: tnsannikova@craocrimea.ru