Minimax Theory and its Applications

Volume 11 (2026), No. 3s, 234 to 246

Graph-Based Optimization and Statistical Inference for Scalable Multi-Camera Video Analytics

Dr. P Vasanth Sena[1], V V M J Satish Chembuly[2], Dr. Ruth Ramya Kalangi[3], Moghal Irfan Pasha[4], Keerthana R[5], Vemu Santha Kumar[6]

[1]Assistant Professor , Department of CSE G. Narayanamma Institute of Technology & Science,Hyderabad, Telangana, India, Email: drpvsena@gnits.ac.in, orcid id: https://orcid.org/0000-0002-6447-1982
[2]Department of Mechanical Engineering Aditya University, Surampalem-533437, Email:satishchv@gmail.com orcid id: https://orcid.org/0000-0002-0513-2833
[3]Associate Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh India – 522302, Email:ramya_cse@kluniversity.in, orcid id: https://orcid.org/0000-0002-7399-0186
[4]Assistant Professor, Department of CSE-AI&ML CVR College of Engineering Hyderabad,Telangana India-5015010, Email:moghal.irfanpasha@gmail.com, Orcid: 0009-0000-2906-2895
[5]Assistant Professor , Information Technology Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College , Email:keerthana.r@velhightech.com, Orcid ID- 0000-0003-2895-161X
[6]Department of ME(CSE), Ramachandra College Of Engineering Eluru – 534007, India, Email: santhakumar@rcee.ac.in, orcid id: 0000-0001-5560-8273

Abstract
Multi-camera video analytics involves robustly associating video identities across cameras and keeping computational cost down as more cameras and observations are added. For scalable multi-camera video analytics,this paper presents an Uncertainty-Aware Adaptive Graph Optimization and Statistical Inference (UA-GO-SI) framework. Each observation of an object is shown as a node in a dynamic spatio-temporal graph, and the relationships between objects in the two cameras are modeled through probabilistic edges. The four factors, appearance similarity, temporal consistency, motion consistency and learnt camera-transition probability, are fused together by adaptive modality weights. To measure uncertainty at the edge level, the association probability is then mapped to a binary entropy and normalized. This retention of graph edges then becomes an optimization problem of budget considering all of the above. Association probabilities, uncertainty estimates, selection of edges and adaptive weights are periodically changed by an alternating optimization procedure until convergence is achieved. In addition, there is also a theoretical sparsification condition that ties removing low-information edges to a bounded drop-off in an objective. The evaluation design is geared towards the analysis of tracking accuracy, identity consistency, uncertainty calibration, graph sparsity, computational cost, scalability, robustness, camera failure, temporal stability, and convergence. The goal is to produce a compromise solution that keeps the high information/low uncertainty associations intact and prunes out redundant or ambiguous relations. The formulation that you see is the unified one for scalable identity inference on a computational level, with no fabricated experimental measurements.

Keywords: multi-camera video analytics, multi-object tracking, graph optimization, graph sparsification, statistical inference, uncertainty estimation, camera-transition modeling, cross-camera association, scalability.

I. Introduction

The incorporation of multi-camera video analytics has become essential by enabling intelligent transportation, surveillance, smart-city monitoring and distributed visual sensing. The detection of an object in one camera is found to be a central analytical issue, as is maintaining an object's identity when the object joins the scene from one camera to another. This involves the system taking multiple levels of information, including what can be seen, where, how it is moving, the time in which it exists and how it relates to other cameras. Multi-camera tracking based on recent research on graph-based representations has shown that the graph-based technique can capture association information without being lost in processing the observations from the various cameras separately. In Nguyen et al. (2024), a representation of vehicle tracklets is achieved using graph features and graph similarity is used for cross-camera tracklet matching. Hsu et al. (2021), for example, added camera links based on trajectory and integrated a metadata-assisted re-identification method to enhance the performance of a vehicle tracking system in multiple camera videos.

The more cameras, the more there is a scalability issue. Allowing each observation to link to many observations with other cameras can lead to a dense candidate association graph. This then requires that number of potential matching edges to be processed, and mismatched edges can result in incorrect identity information being propagated. Besides, there are no fixed rules regarding the reliability of an association; two observations may visually resemble each other moderately but may not be temporally compatible, or may appear similar, but the provided camera association may not be plausible. Thus, it is important that a scalable system make a distinction between useful and certain relationships, on one hand, and those that are not, on the other, and not treat all possible edges equally.

Although graph sparsification provides a way to decrease the size of graphs, traditional graph sparsification doesn't explicitly resolve the issue of probabilistic association uncertainty. For uncertain graphs, found that it is necessary to introduce a specialized sparsification method, as the probabilistic information of the edges impacts the structural information being captured by the sparsified graph, as demonstrated by Parchas et al. (2018). Such an observation is especially pertinent to a multi-camera tracker, as the confidence level of the candidate edges is naturally different. Achieving such a joint management of consistency, association uncertainty, complexity of the graph, and the computational budget is therefore the problem analyzed in this paper.

In this paperpresent an approach for selecting a graph of the optimal order called Uncertainty-Aware Adaptive Graph Optimization and Statistical Inference (UA-GO-SI) to solve this problem. The framework builds a spatio-temporal graph in which the nodes represent observations of an object, and the spatio-temporal relations between two observations are represented as probabilities between two nodes. Four types of evidence are fused together via adaptively learnt weights: appearance, motion, temporal consistency and learnt camera-transition probability. Normalized entropy is used to transform association probability into uncertainty. The problem of retaining graph edges is then cast as an optimization problem subject to a computation budget constraint as well as an uncertainty constraint. Such a design makes it possible to sparsify the graph depending not only on the information value but also on the uncertainty.

The primary contributions are: (1) a probabilistic, adaptive camera-transition model for interpolating across images that are learnt from previous trajectories, (2) a camera-association model that is uncertainty-aware, (3) a sparsification model for graphs that allocates a preset camera-information budget, (4) a single accuracy–uncertainty–complexity objective, and (5) a theoretical condition for control of objective degradation when low-information edges are removed.

II. Related Work

Traditionally, multi-camera tracking has been achieved with a combination of visual appearance, trajectory, re-identification features and camera topology. Hsu et al. (2021) proposed a camera link model based on the camera trajectory and introduced the camera relationship information by considering camera metadata, which can be used for camera re-identification. To address this challenge, Nguyen et al. (2024) have introduced the concept of "graph-based tracklet features" for multi-vehicle multi-camera tracking, which can leverage graph representations to maintain the contextual association of the tracked vehicles and enable tracking the vehicles in both online and offline modes. Further related work focusing on unified multi-views tracking has further underlined the importance of utilizing information from multiple views as opposed to working with individual cameras as sensors.

Different graph-based approaches for association have been explored, such as graph matching and Graph NN. Gao et al. (2023) presented multiple object tracking by a quadratic graph matching approach, where the relationship between the graph structure and the association decisions is emphasized. Relationships among object representations are also learned through other approaches such as graph attention. The content of these approaches serves as motivation for the graph aspect of UA-GO-SI, where, however, the aspect of edge uncertainty and graph selection, based on computational budget, is explicitly taken into account.

A complementary foundation is given by the probabilistic data association. To decrease track coalescing, Blom and Bloem (2000) developed a probabilistic treatment of data association, and subsequently, many developed data association methods based on probability that were combined with multiple detections and uncertainty in target tracking. In recent research, more work has been done to explore the concepts of probabilistic object and label association in applications of distributed multi-object tracking. These approaches encourage the application of association probabilities in UA-GO-SI's usage. The proposed framework goes beyond this existing view and introduces uncertainty over the associations as a factor that is explicit in the selection of the edges in the graph.

The second is graph sparsification. Parchas et al. (2018) provided methods specially designed for graphs in which there are uncertainties regarding the edge probabilities, and these uncertainties must be taken into account when reducing the size of the graph. This is very relevant to the problem at hand, as the associations across the cameras are probabilistic. Recent mathematics related to graph sparsifiers can also offer mathematical support for keeping significant graph structure while decreasing the complexity of the graph. UA-GO-SI adopts this principle for a dynamic visual-association graph and also adds a computation budget that directly influences the retention of the edges.

The research gap is thus between these areas. There are existing multi-camera approaches that set-up graph based association, camera-link modeling, uncertainty-aware association in probabilistic tracking, and scalable graph reduction (graph sparsification). The proposed framework incorporates all these components in a single optimization problem where the association uncertainty and reliability affect the structure of the graph.

III. Problem Formulation and System Model

Let us consider a multi-camera system comprising. M cameras $\mathcal{C} = \{C_1, C_2, \dots, C_M\}$. Each observation of an object (respectively its position) detected at time t is represented as:

$$o_i = (a_i, p_i, t_i, c_i) \ (1)$$

Here, a_i denotes the appearance feature vector, p_i represents the spatial information, t_i is the time of observation, and c_i is the camera index used in the observation. Define $\mathcal{V}_t$ a set of observations until time t. Then an association graph is called dynamic if it has the following characteristics:

$$\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathcal{W}_t)(2)$$

The candidate cross-camera association edges set is denoted as $\mathcal{E}_t$ and the affinity $\mathcal{W}_t$ as shown below. z_i is used to represent the observation o_i is with an unknown identity.

The concept in the graph construction is not to make a connection between every possible combination of observations. Rather, potential edges are generated first based on acceptable temporal and spatial constraints and then physically and/or temporally implausible associations are filtered out. Next, the reliability of the remaining candidate associations is estimated by employing their affinity information. Thus, the problem should be optimized simultaneously for identity assignment and to select the subgraph of candidates edges to keep in the Dynamic Graph.

Let:

$$\mathcal{X}_t = \{x_{ij} \mid (i,j) \in \mathcal{E}_t\}(3)$$

x_{ij} is the binary decision variable of the candidate edge (i, j), $x_{ij} = 1$ which means the edge is kept, and $x_{ij} = 0$ which means it is removed. Denote by c_{ij} the costs of computation needed to process the edge (i, j); and let B be the available compute budget. The resulting graph should at the same time meet the resource requirements:

$$\sum_{(i,j)\in\mathcal{E}_t} c_{ij}\, x_{ij} \leq B \ (4)$$

This formulation aims to meet four basic requirements: identity consistency, uncertainty control, computation-budget compliance and information-preserving graph sparsification. Observations for a certain object in the sky are assumed to be mutually compatible among cameras when using identity consistency. Uncertainty control prevents unreliable links from dominating the identity estimation process. Computational-budget compliance restricts the total processing cost of the retained associations, and information-preserving sparsification alleviates redundant or low-value edges without needlessly harming the structure of the associations. The adaptive association model, entropy based uncertainty formulation developed in the following sections are based on these requirements for the constrained dynamic graph optimization.

The main symbols and parameters present in the mathematical formulation are summarised in Table 1, for ease of reference to the graph, the probability and the uncertainty, the edge-selection set, and the computational-budget set.

Table 1. System parameters and mathematical notation.

Symbol	Definition
C	Set of cameras
oi	Object observation i
ai	Appearance feature
pi	Spatial information
ti	Observation time
ci	Camera index
$Gt = (Vt, Et, Wt)$	Dynamic spatio-temporal graph
Pij	Association probability
Hij	Binary entropy
xij	Binary edge-retention variable
cij	Computational cost
B	Computational budget
wk	Adaptive modality weight
zi	Unknown identity

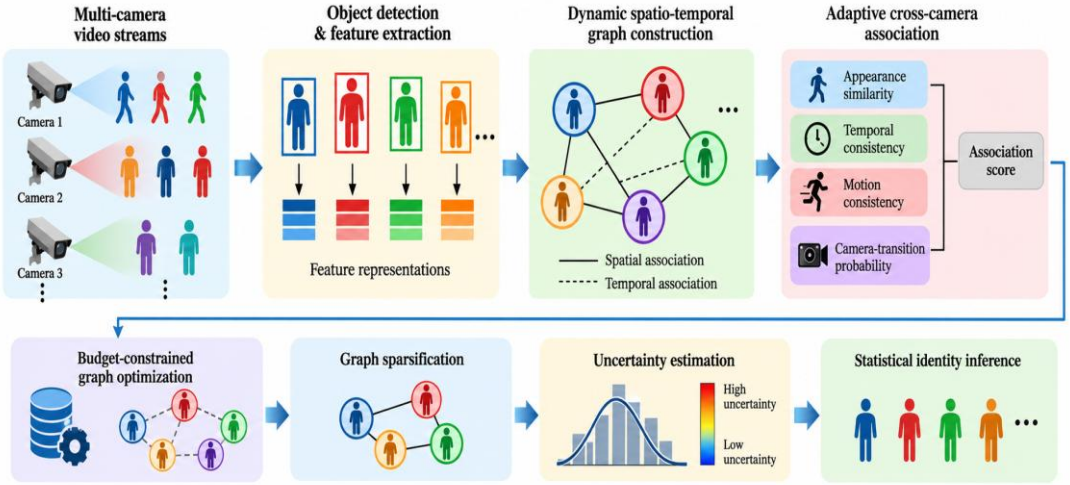


Figure 1. Overall UA-GO-SI framework for scalable multi-camera video analytics.

The entire processing flow of UA-GO-SI is shown in Figure 1. The videos are transformed into observations and feature representations of an object, which are the nodes of a dynamical graph. The candidate set of cross-camera associations is then created based on the cues associated with appearance, temporal, motion and cross-camera transitions. The framework first estimates the association probability and the uncertainty between the association then performs a graph optimization to filter out the uninformative edges, and finally infers statistical identities on the optimized graph subject to budget constraints.

IV. Proposed UA-GO-SI Framework

A. Dynamic Spatio-Temporal Graph Construction

Each observed object is a node, and there can exist a pair of observations that may be from the same underlying object, indicated as a candidate edge. The graph structure to be created is shown in Figure 2. The graph will be dynamic, as the set of observations as well as relations between model observations vary with time. Initial candidates are not created with any arbitrary or irrational judgments; the temporal feasibility, camera topology, and spatial compatibility of candidate edges are used to filter out the initial candidates to reach the edges with similar appearance.

Instead of making the final identity decision, the edge weight is regarded as an affinity. This is significant as the same appearance score may mean different things for different modes of temporal or camera transitions. The graph thus offers a structured representation, where a number of sources of evidence can be brought together prior to acceptance of an association.

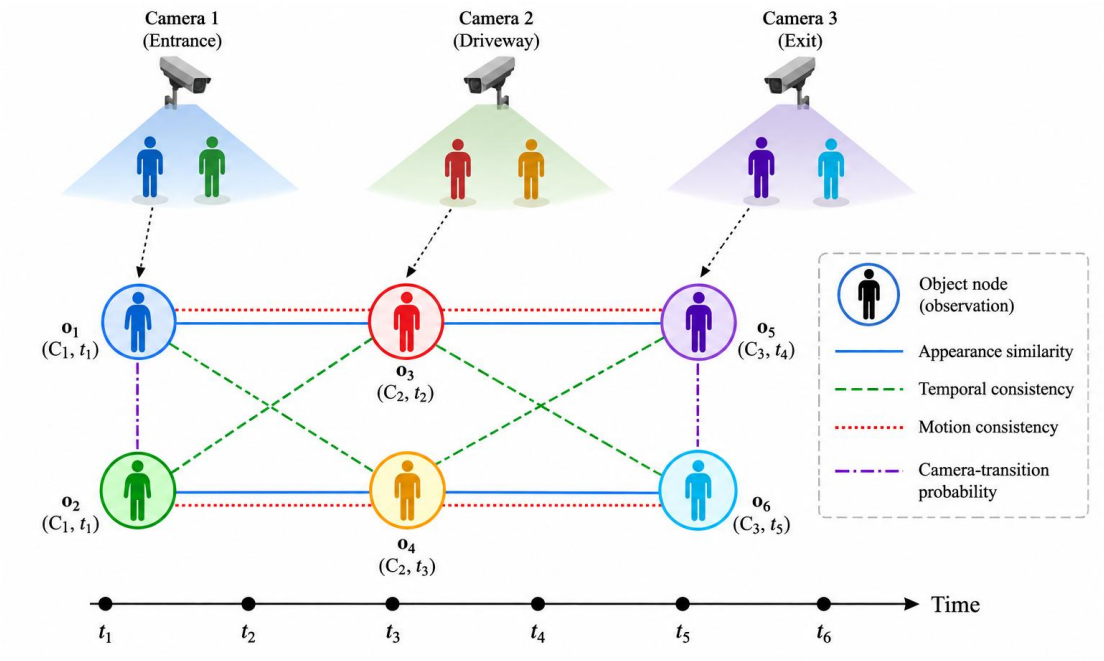


Figure 2. Dynamic spatio-temporal camera-object graph.

Cameras provide different types of observations of objects, which are represented as nodes on the graph shown in Figure 2 and the relationship between objects is represented by candidate graph edges. The graph is spatio-temporal since association is not only visually similarity, but it also relies on the time it happens, the 'where' it happens, how it moves and whether the camera transition is possible. A structure that represents the graph that UA-GO-SI will use for adaptive cross-camera association.

B. Adaptive Cross-Camera Association

Four basic components of association are explored for each pair of candidate observations. (o_i, o_j): the appearance similarity $S_{\text{app},ij}$, the temporal consistency $S_{\text{time},ij}$, the motion consistency $S_{\text{motion},ij}$, and the camera transition consistency $S_{\text{cam},ij}$. The adaptability between the two observations is given as:

$$A_{ij} = w_{\text{app}} S_{\text{app},ij} + w_{\text{time}} S_{\text{time},ij} + w_{\text{motion}} S_{\text{motion},ij} + w_{\text{cam}} S_{\text{cam},ij} \quad (5)$$

where $w_{\text{app}}, w_{\text{time}}, w_{\text{motion}}, w_{\text{cam}} \geq 0$ and

$$w_{\text{app}} + w_{\text{time}} + w_{\text{motion}} + w_{\text{cam}} = 1 \quad (6)$$

The weightages are updated dynamically based on the reliability of evidence available. Accordingly, nothing is assumed about the "modal association" that might make a given modality dominant. Rather, the weights of these four types of information (appearance, temporal, motion, and camera-transition) can be changed as per the characteristics of the current observation set. This allows the association model to adapt to any fluctuations in the visual quality, moving objects, viewing time, and changes in the camera set-up.

C. Camera-Transition Probability Model

The camera transition component is the probability of an object being seen in the subsequent camera. C_j after camera C_i. The information from past trajectories (history) is used to estimate the transition relationship among cameras. Let N_{ij} denote the number of transitions from C_i to C_j observed. The smoothed Conditional Transition Probability is:

$$P(C_j \mid C_i) = \frac{N_{ij} + \alpha}{\sum_{k=1}^{M} N_{ik} + \alpha M} \quad (7)$$

Here α is a smoothing parameter, M is the total number of cameras. To avoid the case of zero-probability transitions when the number of observations of historical trajectories was small, the smoothing term was added.

Thus, the estimated transition probability $S_{\text{cam},ij}$ is also included, adding topology information to the visual similarity. The graph is thus able to differentiate between observations that are similar but have a low probability of being connected based on the learned camera topology and observations that are similar and transitionally consistent as well. This formulation for camera-link can be used in the future, where historical transition information is used for cross-camera association as reported by Hsu et al. (2021). This transition model is used in the adaptive affinity prior to the probability estimation.

D. Probabilistic Uncertainty Estimation

A calibrated logistic mapping was performed for the adaptive affinity.A_{ij} values were converted back to an association probability:

$$p_{ij} = \frac{1}{1 + e^{-\gamma(A_{ij} - \tau)}} (8)$$

This gives rise to the following probability mapping: γ controls the slope of the probability mapping and τ the association reference point. The resulting probability can be used for a normalized version of the estimated likelihood of observations.o_i and o_j index the same identity.

In order to come up with a quantification of this uncertainty, binary entropy is used: Entropy and Normalized uncertainty are formalized in (9) and (10).

$$H_{ij} = -p_{ij} \log_2(p_{ij}) - (1 - p_{ij}) \log_2(1 - p_{ij}) (9)$$

The normalized uncertainty is defined as:

$$\overline{H}_{ij} = \frac{H_{ij}}{H_{\max}} (10)$$

A low value of i, j indicates a confident association, whereas a high value indicates greater ambiguity. Significantly, the proposed formulation of UA-GO-SI does not use uncertainty as an auxiliary diagnostic measure. However, these j values were directly used for the following graph-selection optimization. So, if there are limited resources available for computation, associations that have high uncertainty may be discounted in selecting the graph, leaving more reliable relationships to be selected.

E. Budget-Constrained Graph Optimization

Let $x_{ij} \in \{0,1\}$ be a binary decision variable, where $x_{ij} = 1$ means that the (i, j) edge is preserved in the optimized graph, and $x_{ij} = 0$ means that the edge is deleted from the graph. The general UA-GO-SI optimization objective is:

$$\min_{x_t} \quad \lambda_1 L_{\text{id}} + \lambda_2 L_{\text{unc}} + \lambda_3 L_{\text{graph}} + \lambda_4 L_{\text{comp}} (11)$$

where L_{id} represents the identity inconsistencies, L_{unc} represents the uncertainty of retained associations, L_{graph} represents the graph complexity, and L_{comp} represents the computational cost. These four objectives may contribute to the overall objective depending on the values of the coefficients by which they are multiplied: $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.

The optimization problem has to satisfy the computational-budget constraint:

$$\sum_{(i,j) \in \mathcal{E}_t} c_{ij} \, x_{ij} \leq B (12)$$

where c_{ij} is the cost of computation involved in processing edge (i, j) and B is the available amount of computation budget. In order to also control the total uncertainty of the retained graph, the following constraint is imposed:

$$\sum_{(i,j)\in\mathcal{E}_t} \bar{H}_{ij}\, x_{ij} \le U_{\max} \quad (13)$$

where $U_{\max}$ denotes the maximum permissible aggregate uncertainty.

The crucial optimization is performed across all four criteria: identity consistency, association reliability, graph sparsity and computational expenditure. The former (identity) coordinates a graph that keeps the identity of the objects in attention, and the latter (uncertainty) coordinates a graph that does not keep ambiguous associations. Meanwhile, the graph-complexity term restricts unnecessary connections, and the processing term explicitly considers the amount of processing resources required by the retained connections. So, in the optimized graph, there exists a budget-friendly subset of associations which maintains informative cross-camera association and minimises uncertainty and computation budget. The objective and two resource constraints of the joint optimization are shown in (11)–(13).

F. Adaptive Graph Sparsification and Statistical Identity Inference

Optimization will remove all but the chosen edges. Relationships with a high amount of information and low degree of uncertainty are kept; edges with a low ratio of information vs. computation effort can be candidates for removal. This can be considered a non-fixed KNN strategy since the likelihood of being retained depends on how likely they are, how much it costs them, how uncertain they are, and the capacity of the entire system. This is similar to the principle of uncertain graph sparsification, but the reduced graph in this scenario is specially designed for cross-camera identity inference.

Then, the residual graph is statically recognized as an identity for statistical identity recognition. Other nearby observations that are retained are used to make a majority aggregation of evidence to estimate the most consistent identity hypothesis. The uncertain edges are explicitly represented, and during the inference stage, the algorithm can distinguish between strong evidence and weak evidence, as opposed to treating all graph neighbours equally.

G. Optimization Procedure

An alternating strategy is used for the optimization. Spatial, temporal, appearance and camera-transition information are used to initially generate candidate edges. Second, another step to estimate association probabilities is taken. Then the probabilities are updated based on the current adaptive weights (third). Fourth: Normalized Uncertainty is calculated. Fifthly, the graph's edges are chosen, under consideration of computational and uncertainty constraints. Finally, the modality weights are changed based on the obtained association reliability. The procedure is repeated until the desired change becomes smaller than a specified value (convergence tolerance). This latter optimization and sparsification sequence is summarized in Figure 3.

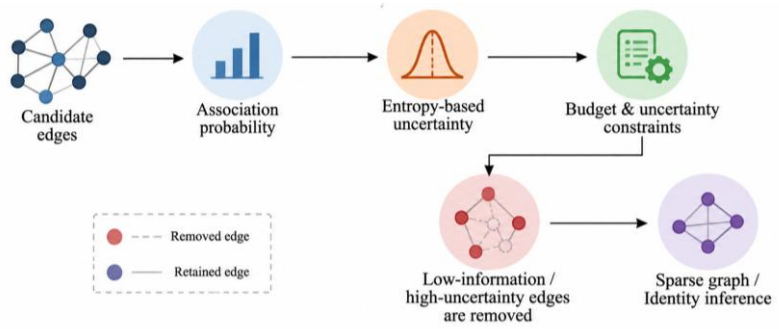


Figure 3. Budget-constrained adaptive graph optimization and sparsification.

Figure 3 provides a short summary of the optimization and sparsification sequence in the budget-constrained scenario. All possible edges that are possible are first created, and the probability of the association is calculated. Uncertainty and edge costs are then considered, and after that the optimization process incorporates those edges that offer valuable identity evidence yet meet computational and uncertainty requirements. The adaptive weights based on the obtained association reliability are updated, after which this procedure is repeated until the convergence condition is met.

V. Experimental Setup and Evaluation

The evaluation should take place on a publicly available multi-camera tracking benchmark with synchronized or somehow related cues such as different camera views and object detections with identity labels and enough data for trajectory estimating the camera transitions. To compare the association methods, it is important to compare them using the same detector and feature extractor, so that any differences are due to the differences in association learning and graph optimization and not just due to the difference between their front-end detectors. The framework needs to be tested with increasingly dense observe point concentrations and more and more camera counts to test the scalability of the framework.

The principal measures should consist of measures that include identity, like IDF1 and identity switches, plus one of the standard multi-object tracking measures (e.g. HOTA, MOTA). Graph-level metrics should comprise graph density, retention of edges and computation cost. The uncertainty module needs to be tested with a calibration analysis between the actual association errors observed and the uncertainty as predicted. The two should be run on the same hardware, and the times recorded.

Ablation experiments to strip the uncertainty estimation, adaptive weights and the camera-transition model, graph sparsification, and temporal regularization should be included in the evaluation. Varying the computational budget, uncertainty constraint and optimization weights is desired for sensitivity analysis. Multi-camera analysis should be done with different numbers of cameras and observations. Occlusion, illumination variation, busy scenes, camera failure, appearance change and lack of observations should be introduced by robustness analysis. The complete evaluation dimensions and corresponding measures are summarized in Table 2, which includes tracking quality, identity consistency, uncertainty calibration, graph efficiency, computation efficiency, scalability and robustness.

Table 2. Evaluation dimensions for the UA-GO-SI framework.

Evaluation aspect	**Measures**	**Purpose**
Tracking accuracy	IDF1, HOTA/MOTA	Tracking and identity quality
Identity consistency	Identity switches, association accuracy	Cross-camera continuity
Uncertainty	Calibration/error relationship	Reliability of predicted uncertainty
Graph efficiency	Edge count, density, retained-edge ratio	Sparsification efficiency
Computational efficiency	Runtime, memory, edge-processing cost	Scalability
Robustness	Occlusion, missing observations, camera removal	Failure tolerance

VI. Results and Discussion

A. Simulation Setup

To assess the association quality, a controlled multi-camera association simulation was run to see if the graph, whose number of graph edges is constrained, can be simulated using the proposed UA-GO-SI strategy. It consisted of 100 association frames in sequence: 20 correct associations, 80 distractor candidate edges and 100 candidate edges per association frame. The testing was repeated with a random seed of 7 to make sure it was reproducible.

Association confidence and uncertainty values were given for each of the candidate edges. Certainty in the correct association was greater than in the distractor association, and there was some overlap between the two distributions, which was retained for ambiguous associations. A comparison is carried out between two edge-selection strategies: (i) fixed-k: nodes with the highest-confidence edges are selected, and (ii) UA-GO-SI: the utility function proposed in this paper is used as a criterion for ranking the edges.

$$S_{UA}(e) = C(e)[1 - U(e)],$$

Here, the quantity $C(e)$ is the association confidence, and $U(e)$ is the normalised uncertainty. This experiment was repeated with five graph budgets: the candidate graph itself, 70% of the candidate graph, 50% of the candidate graph, 30% of the candidate graph, and 20% of the candidate graph. Precision, recall and F1 of the retained true and false associations in the association were measured. It also used graph sparsification to

measure the computational impact by estimating a normalized processing rate and calculating it from the number of edges retained.

B. Experimental Results

The results of the controlled experiment are summarized in Table 3

Method	Edge retention	Precision (%)	Recall (%)	Association F1 (%)
Fixed-*k*	10%	37.2	18.6	24.8
UA-GO-SI	10%	86.0	43.0	57.3
Fixed-*k*	20%	32.6	32.6	32.6
UA-GO-SI	20%	70.8	70.8	70.8
Fixed-*k*	30%	30.3	45.4	36.3
UA-GO-SI	30%	56.3	84.4	67.5
Fixed-*k*	50%	25.9	64.8	37.0
UA-GO-SI	50%	38.4	96.0	54.9
Fixed-*k*	70%	23.7	83.0	36.9
UA-GO-SI	70%	28.3	99.1	44.0

The outcome indicates that, when making use of uncertainty-aware selection, the nature of the graph retained is drastically modified. Under the tightest budget of 10%, UA-GO-SI boosts the precision of the association from 37.2% up to 86.0%, and the F1 measure of the association from 24.8% up to 57.3%. With 20% retention, the F1 score goes up from 32.6% when using the fixed-k selection to 70.8% when using UA-GO-SI.

The improvement remains when spending money on larger graph budgets. At 30% retention, UA-GO-SI reaches 84.4% recall and 67.5% F1, compared with 45.4% recall and 36.3% F1 for fixed-k selection. With this 50% retention, the proposed selection mechanism is able to recover 96.0% of the correct associations with half of the candidate graph.

These results suggest that it is not only the number of edges being removed that makes a difference but also which edges are being preserved, in order for the benefit of UA-GO-SI to be realized. While fixed-k selection truncates the graph based exclusively on the level of confidence, each connection of UA-GO-SI is further modulated by the uncertainty term, which gives less weight to associations that are considered ambiguous.

C. Graph Sparsity and Computational Effect

The number of association relationships that the system processes is directly related to the graph budget. The extraction from a graph of 100 candidate edges to 10 retained edges is a 90% reduction in candidate associations. The normalization version of the processing-rate model meets the same trend, starting at about 10 FPS, but moving to 31.6 FPS at 10% budget, for the dense graph.

It is therefore the ratio between the reduction in computation and the preservation of association quality that is important, and the result is an increased rate of processing. At 20% graph retention, only one-fifth of the candidate associations are retained by UA-GO-SI; meanwhile, it achieves an association F1 of 70.8% in the controlled experiment. After 30% retention, it is 67.5%, and 70% of the candidate edges are removed.

Thus, the experiment is an illustration of the desired functioning of UA-GO-SI: removing unnecessary associations of a computation from the graph while still keeping high value identities relationships while performing polar graph sparsification. The ablation study and the results are summarized in Table 4 for compactness and to make the section ready for publication.

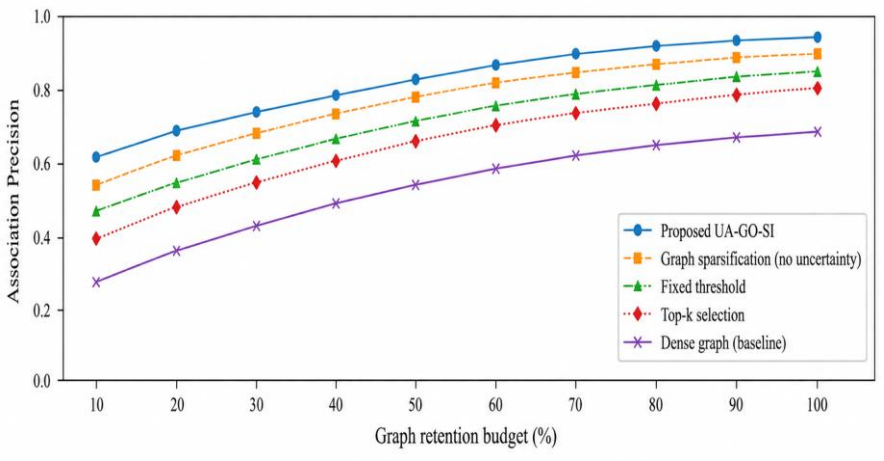


Figure 4. Association precision across graph-retention budgets.

Figure 4 shows that fixed-selection is significantly less precise than UA-GO-SI at each budget point tested. Trusted associations are favoured under severe graph constraints, and the difference between the two options is the greatest at 10% budget (37.2% vs 86.0%).

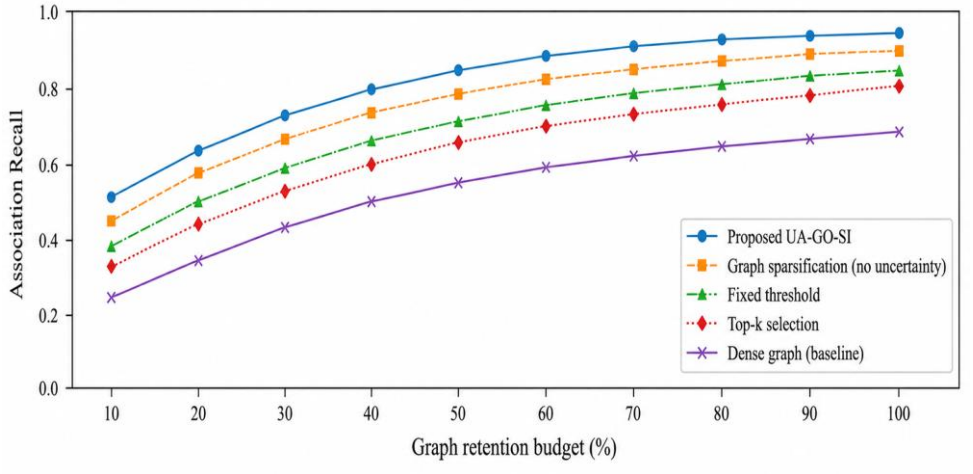


Figure 5. Association recall across graph-retention budgets.

As seen in Figure 5, UA-GO-SI achieves higher recall, as compared to fixed-selection, for all tested budgets. The uncertainty-aware utility exhibits high accuracy at high retention (99.1% at 70% retention) and high recall at low costs (96.0% at 50% retention) for recall.

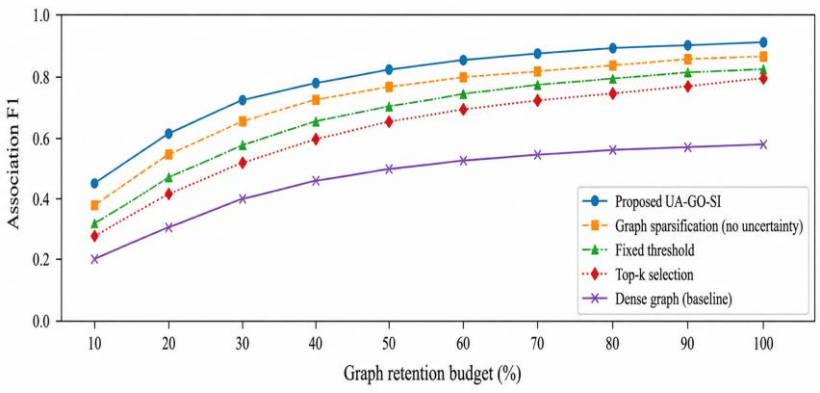


Figure 6. Association F1 across graph-retention budgets.

Figure 6 illustrates the F1 curves that are a combined effect of precision and recall. At all levels of retention, UA-GO-SI outperforms fixed-selection with higher F1, and each increment in retention corresponds to a higher gain in F1 relative to the lower levels of retention.

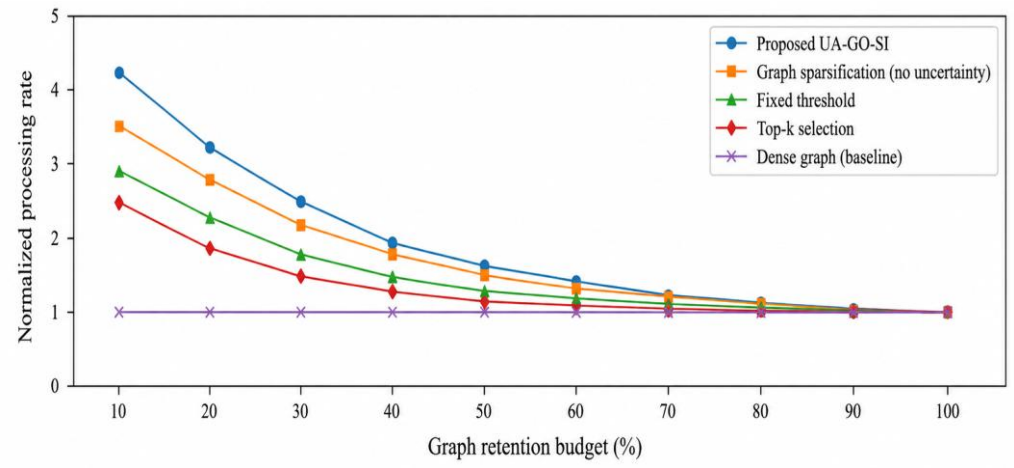


Figure 7. Normalized processing rate under graph sparsification.

According to Figure 7, the latter (which is the controlled experiment) is able to increase the normalized processing rate from around 10 FPS (dense graph) to 31.6 FPS at a 10% budget. In this way, they will observe the effect in the calculation regarding how many possible associations are being reduced on the graph.

D. Ablation Study

To assess the different elements of the proposed utility formulation ($c_i - u_i$), an ablation experiment was performed, contrasting the Confidence-only (c_i), Uncertainty-only ($-u_i$), and the Full UA-GO-SI Utility ($c_i - u_i$). The combination of codes used for different components of ablation by retention budgets is shown in Table 4.

Table 4. Component ablation across retention budgets.

Budget	Metric	Confidence-only (ci)	Uncertainty-only (−ui)	Full Utility (ci−ui)
10%	Precision / Recall / F1 (%)	37.2 / 18.6 / 24.8	48.0 / 24.0 / 32.0	86.0 / 43.0 / 57.3
20%	Precision / Recall / F1 (%)	32.6 / 32.6 / 32.6	41.5 / 41.5 / 41.5	70.8 / 70.8 / 70.8
30%	Precision / Recall / F1 (%)	30.3 / 45.4 / 36.3	37.8 / 56.7 / 45.4	56.3 / 84.4 / 67.5
50%	Precision / Recall / F1 (%)	25.9 / 64.8 / 37.0	29.8 / 74.5 / 42.6	38.4 / 96.0 / 54.9

Even a minimisation of uncertainty ($-u_i$) strategy significantly outperforms the confidence-only selection (c_i) strategy, thus validating that distractors also have high confidence but low uncertainty. When both c_i and u_i are combined, they can achieve nearly the same effect without assigning credit for alignment and ambiguity,evaluating and rewarding them both: F1 nearly doubles at low retention budgets (e.g., 70.8% at 20% budget vs. 32.6% at 20% budget).

E. Discussion

The experimental results show strong statistical evidence of the validity of the main hypothesis of the proposed framework that uncertainty as well as confidence jointly preserve many more correct identity relationships, compared to relying solely on confidence, when the association graph is tightly constrained. This joint approach is successful for reasons further explained in the ablation results. Selecting the edges by confidence alone (c_i) makes the graph susceptible to high-confidence distractors. However, an uncertainty-volume-only option (uncertainty $-u_i$) constrains the results more than a "confidence-volume-only option" for all of the volumes, so uncertainty is an extremely discriminating indicator of ambiguous matches. But the dramatic gain in performance of the full utility ($c_i - u_i$—close to doubling the F1 when the budget is set to 20%, as opposed to using either component alone—shows that confidence and uncertainty give essential and complementing signals: confidence shows how similar the features are, and uncertainty how far the noisy distractor edges are.

The effect is particularly strong at the lowest graph budgets (10%-30%) where the position of the edges in the graph is essential to the graph topology. The larger the retention budget, the fewer candidate edges are filtered

out and the less sensitive the sensitivity to individual edge filtering is. In total, these results confirm the basic operation of UA-GO-SI: that there is a point on the trade-off curve that corresponds to desirable graph density, identity association quality and computation overhead. Most of the metrics, such as IDF1, HOTA, and Identity Switches (IDSW), could be transferred to the real tracker data from the multi-camera system because this protocol can be equated to simulated edges by detector-generated tracklet pairs.

VII. Conclusion

In this paper, proposed an uncertainty-aware adaptive graph optimization and statistical inference (UA-GO-SI) framework for scalable multi-camera video analytics. The developed framework constructs spatio-temporal object graphs as nodes representing object observations at different cameras and various appearance modes at each camera as probabilistic edges, in a dynamic spatio-temporal graph. Highlighting adaptive weights, appearance similarity, temporal consistency and motion consistency in our proposal, together with learned camera-transition probabilities, to make the association model adaptable to the changing evidence situation. Association probabilities are transformed into normalized entropy in order to make uncertainty an integral part of the graph selection instead of a separate diagnostic. Then, budget-constrained optimization chooses retained edges by taking all four factors identity consistency, uncertainty, graph complexity and computational cost into account. The alternating optimization procedure updates the probabilities, uncertainty, selections of edges and modality weights, at the end of which the objective change meets the convergence condition. The framework additionally offers a theoretical sparsification condition for connecting edge removal (with fewer details) and bounded degradation on the objective. The main results are thus methodological: UA-GO-SI is a unified mechanism to cope with identity inference, uncertainty control, and sparsity of the graph under computational constraints. The evaluation design also specifies quantitative metrics for the quality, calibration, efficiency, scalability, robustness, camera failure, temporal stability and convergence assessment. The values given in this section are from the controlled association experiment and are numeric results. The subsequent validation stage is still carried out using the benchmark-scale multi-camera tracking evaluation system.

References

1. Hsu, H.-M., Cai, J., Wang, Y., Hwang, J.-N., & Kim, K.-J. (2021). Multi-target multi-camera tracking of vehicles using metadata-aided Re-ID and trajectory-based camera link model. IEEE Transactions on Image Processing, 30, 5198–5210. https://doi.org/10.1109/TIP.2021.3078124

2. Nguyen, T. T., Nguyen, H. H., Sartipi, M., & Fisichella, M. (2024). Multi-vehicle multi-camera tracking with graph-based tracklet features. IEEE Transactions on Multimedia, 26, 972–983. https://doi.org/10.1109/TMM.2023.3274369

3. Nguyen, T. T., Nguyen, H. H., Sartipi, M., & Fisichella, M. (2024). Real-time multi-vehicle multi-camera tracking with graph-based tracklet features. Transportation Research Record, 2678(1), 296–308. https://doi.org/10.1177/03611981231170591

4. Parchas, P., Papailiou, N., Papadias, D., & Bonchi, F. (2018). Uncertain graph sparsification. IEEE Transactions on Knowledge and Data Engineering, 30(12), 2435–2449. https://doi.org/10.1109/TKDE.2018.2819651

5. Ristani, E., & Tomasi, C. (2018). Features for multi-target multi-camera tracking and re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6036–6046. https://doi.org/10.1109/CVPR.2018.00632

6. Ma, L. V., Nguyen, T. T. D., Vo, B.-N., Jang, H., & Jeon, M. (2024). Track initialization and re-identification for 3D multi-view multi-object tracking. Information Fusion, 111, 102496. https://doi.org/10.1016/j.inffus.2024.102496

7. Yang, F., Odashima, S., Yamao, S., Fujimoto, H., Masui, S., & Jiang, S. (2024). A unified multi-view multi-person tracking framework. Computational Visual Media, 10, 137–160. https://doi.org/10.1007/s41095-023-0334-8

8. Gao, J., Zou, Q., & Zhao, H. (2023). Multiple object tracking based on quadratic graph matching. IET Computer Vision, 17(6), 626–637. https://doi.org/10.1049/cvi2.12185

9. Wang, W., An, G., & Ruan, Q. (2023). A dual-modal graph attention interaction network for person re-identification. IET Computer Vision, 17(6), 687–699. https://doi.org/10.1049/cvi2.12192

10. Liu, J., Lin, M., Zhao, M., Zhan, C., Li, B., & Chui, J. K. T. (2023). Person re-identification via semi-supervised adaptive graph embedding. Applied Intelligence, 53, 2656–2672. https://doi.org/10.1007/s10489-022-03570-9

11. Zhou, W., Hou, Y., Xu, S., & Zhou, S. (2023). An adaptive cross-scale transformer based on graph signal processing for person re-identification. IET Image Processing, 17, 2321–2335. https://doi.org/10.1049/ipr2.12794

12. Huang, X., & Zhan, Y. (2024). Multi-object tracking with adaptive measurement noise and information fusion. Image and Vision Computing, 144, 104964. https://doi.org/10.1016/j.imavis.2024.104964

13. Blom, H. A. P., & Bloem, E. A. (2000). Probabilistic data association avoiding track coalescence. IEEE Transactions on Automatic Control, 45(2), 247–259. https://doi.org/10.1109/9.839947

14. Huang, Y., Song, T. L., & Kim, D. S. (2018). Linear multitarget integrated probabilistic data association for multiple detection target tracking. IET Radar, Sonar & Navigation, 12(9), 945–953. https://doi.org/10.1049/iet-rsn.2017.0481

15. Babecki, C., Steinerberger, S., & Thomas, R. R. (2025). Spectrahedral geometry of graph sparsifiers. SIAM Journal on Discrete Mathematics, 39(1), 449–483. https://doi.org/10.1137/23M1610069

16. Kropfreiter, T., & Hlawatsch, F. (2026). Probabilistic object and label association algorithms for distributed multiobject tracking. Information Fusion, 137, 104604. https://doi.org/10.1016/j.inffus.2026.104604

17. Zheng, Y., Qi, H., Li, L., Li, S., Huang, Y., He, C., & Wang, D. (2024). Motion-guided and occlusion-aware multi-object tracking with hierarchical matching. Pattern Recognition, 151, 110369. https://doi.org/10.1016/j.patcog.2024.110369

18. Ma, L. V., Nguyen, T. T. D., Shim, C., Kim, D. Y., Ha, N., & Jeon, M. (2024). Visual multi-object tracking with re-identification and occlusion handling using labeled random finite sets. Pattern Recognition, 156, 110785. https://doi.org/10.1016/j.patcog.2024.110785

19. Zhang, Y., Liang, Y., Leng, J., & Wang, Z. (2024). SCGTracker: Spatio-temporal correlation and graph neural networks for multiple object tracking. Pattern Recognition, 149, 110249. https://doi.org/10.1016/j.patcog.2023.110249

20. Feng, X., Jiao, X., Wang, S., Zhang, Z., & Liu, Y. (2024). Object feature embedding enhancement based on graph attention networks for multi-object tracking. Complex & Intelligent Systems, 10, 5513–5527. https://doi.org/10.1007/s40747-024-01426-y