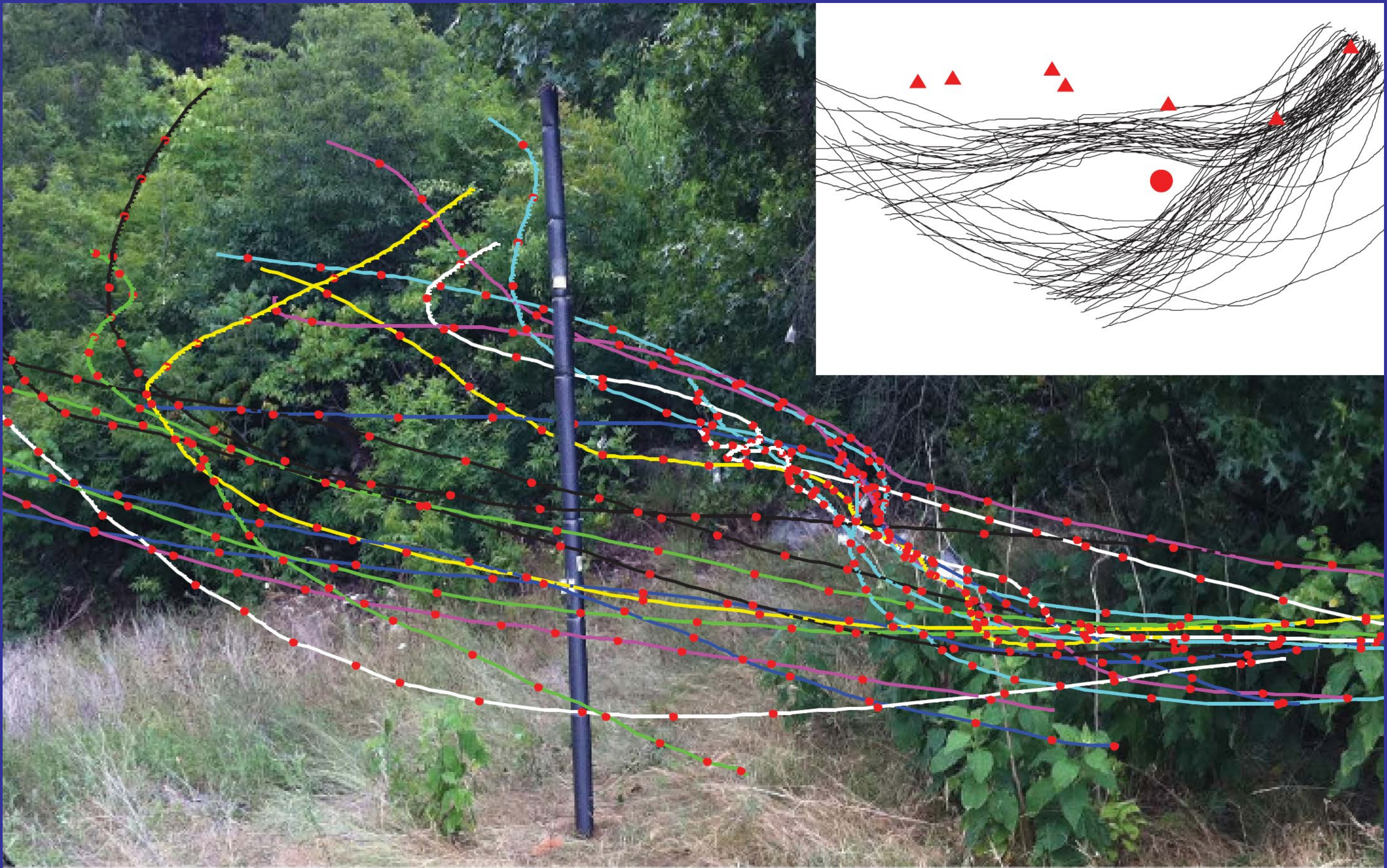
CS 585 Lecture on Multi-View Multi-Object Tracking

Margrit Betke
Computer Science Department
Boston University

April 23, 2024

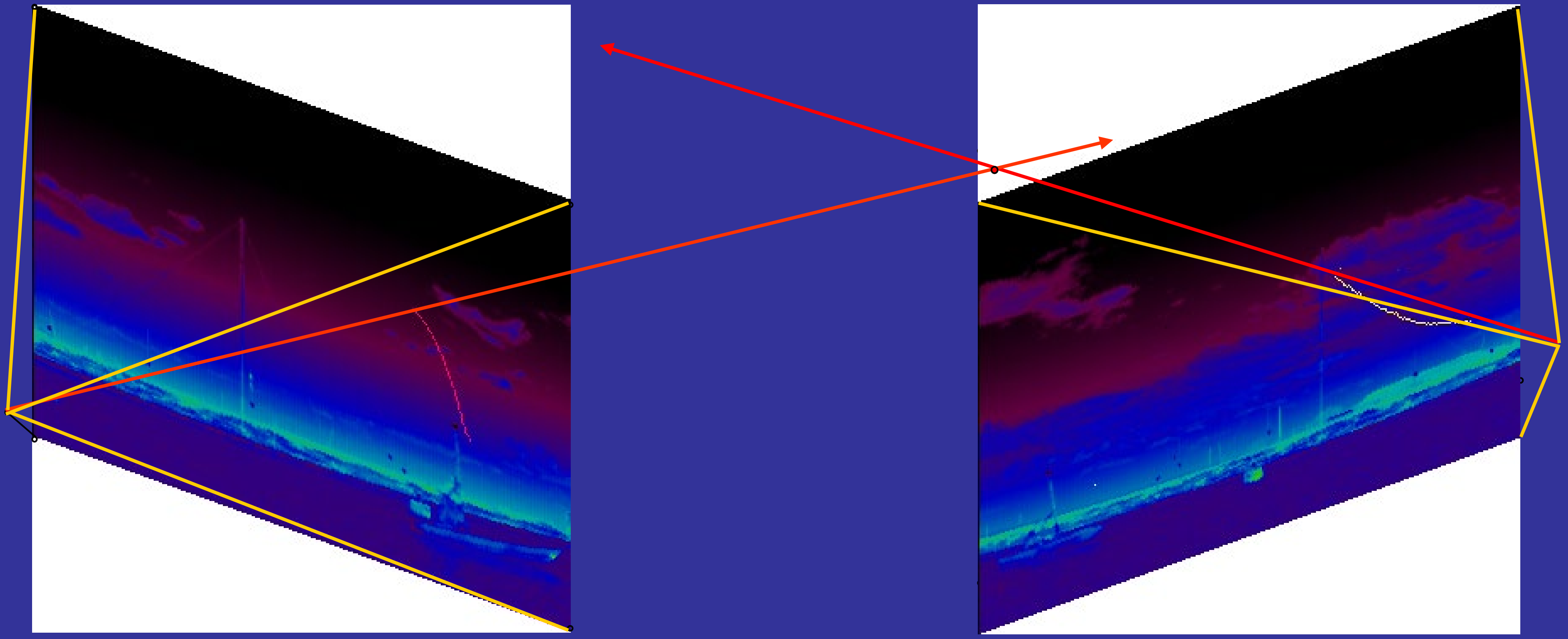
Opportunities for Studying Wildlife

How do bats fly with respect to their environment?



Stereoscopic 3D Reconstruction via Triangulation

3D position of single object



left camera

right camera

Tracking Multiple Objects via Two Data Association Methods

Previous Class:

- Cluster-based Approach: uses Hungarian method to match measurements and objects in cluster.
- Greedy Approach: “greedily” favors objects with long observation histories. Matching process is started by matching longest-observed object and its nearest measurement. Then second-longest observed object in cluster is matched with its nearest measurement, etc.

- We learned methods to track multiple objects
- We learned methods to estimate the 3D position of an object seen in multiple views

Multiview 2D measurements -> 3D estimate

Multiple moving objects in 2D -> 2D tracks

- We learned methods to track multiple objects
- We learned methods to estimate the 3D position of an object seen in multiple views

Multiview 2D measurements -> 3D estimate

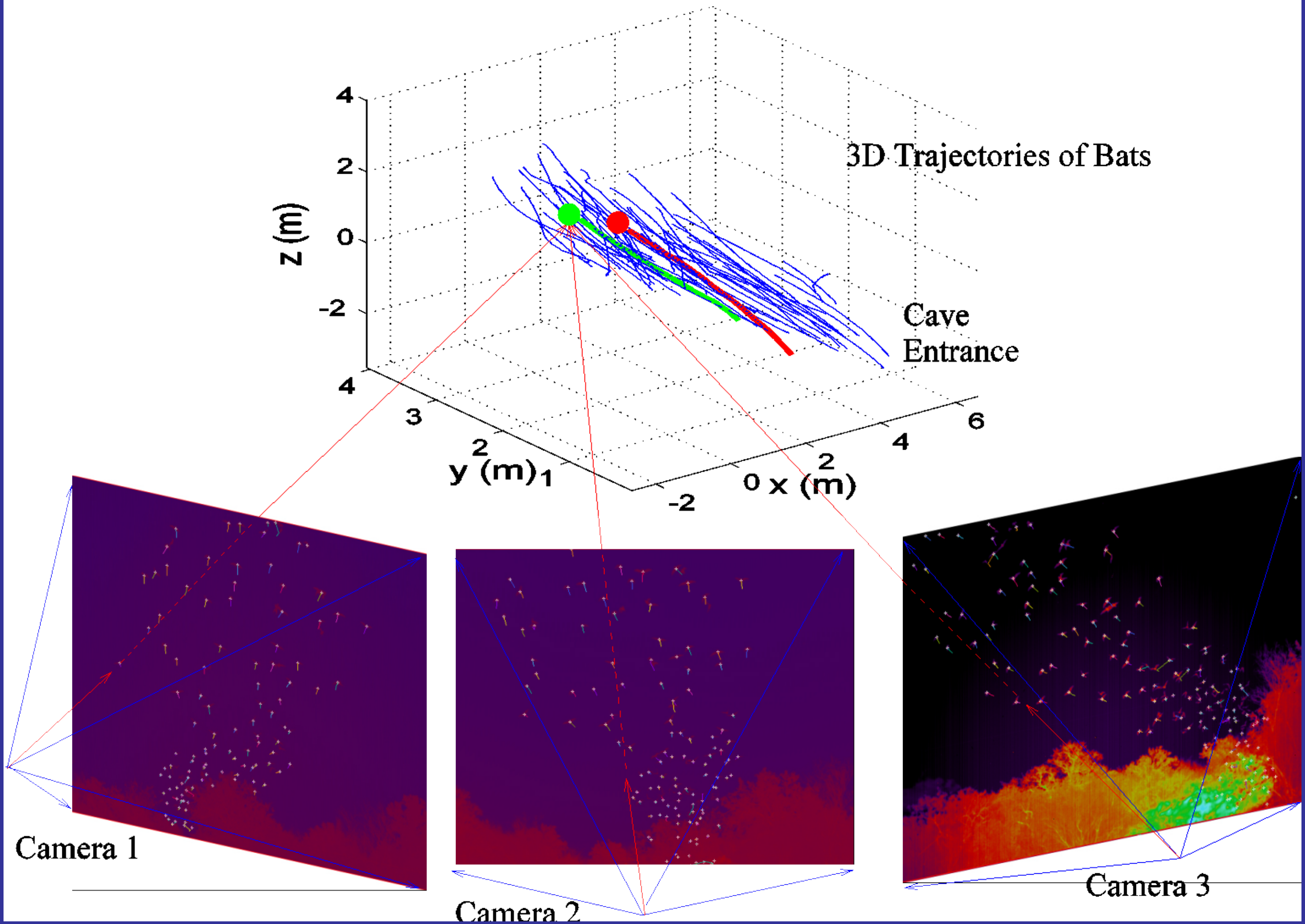
Multiple moving objects in 2D -> 2D tracks

- What about tracking multiple objects seen in multiple views?

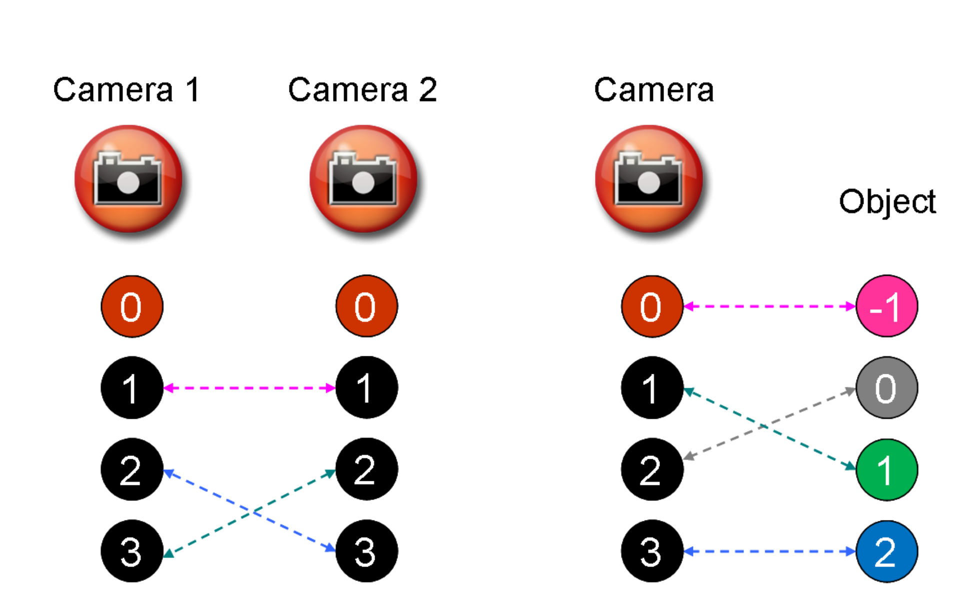
  Sequence of multiview 2D measurements

  -> Multiple 3D object tracks

3D Trajectories of Multiple Objects from Multiple Views, i.e., Multiple Cameras

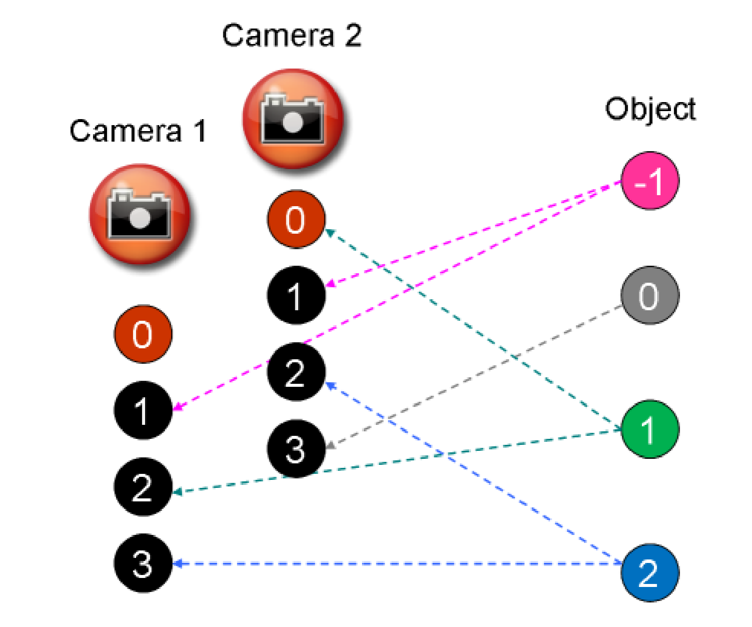


Data Association

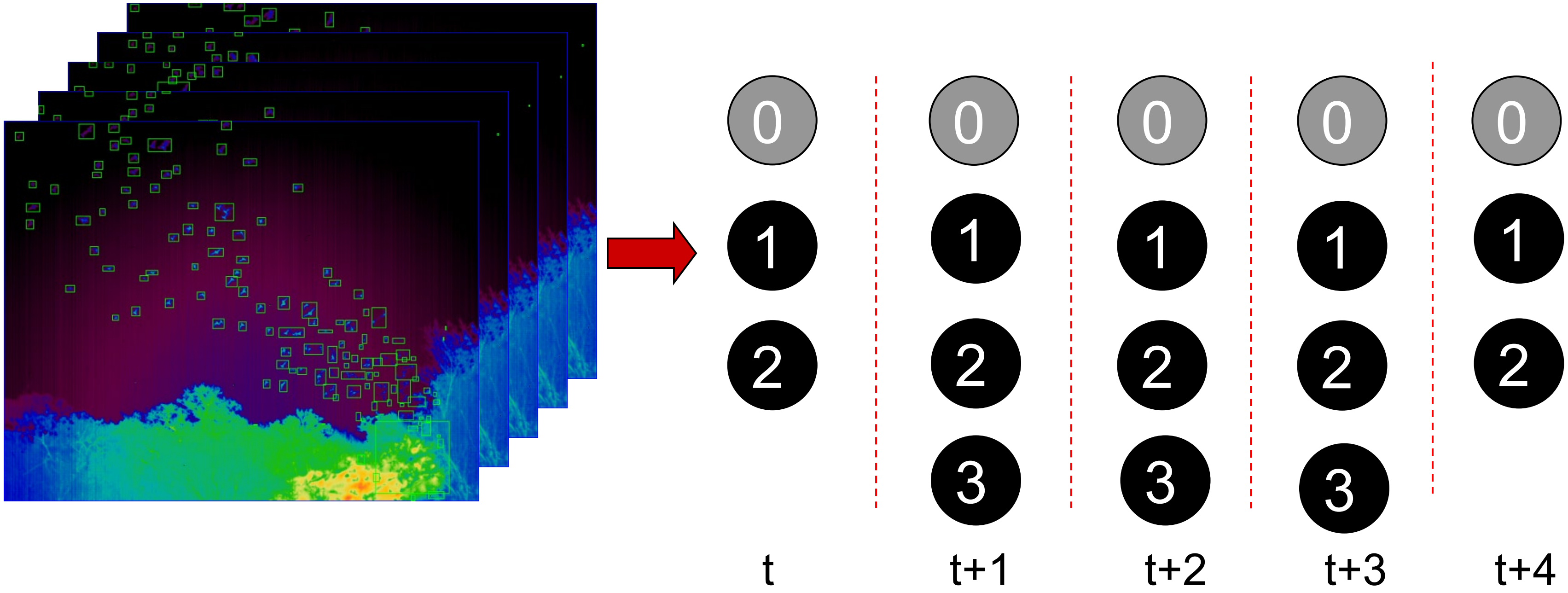


Spatial Association

Temporal Association

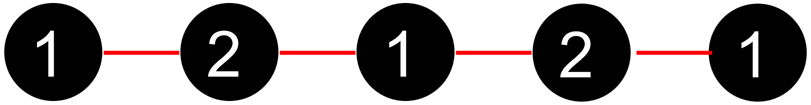


Data Association Problem



2 z_{t,i_t} Measurement / Observation

0 $z_{t,0}$ Missing detection

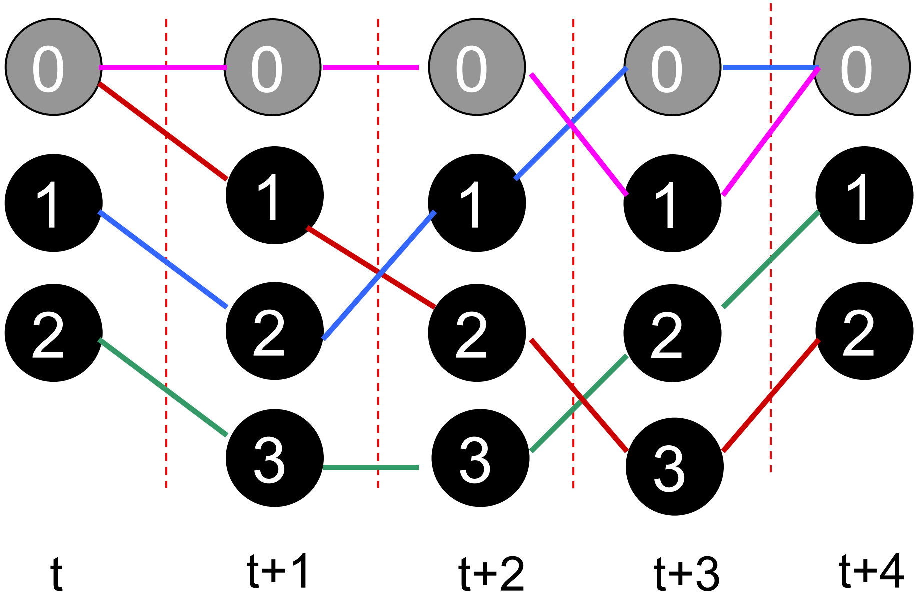


Track τ_i with cost $c_{i_1 i_2 i_3 i_4 i_5}$

$$Z_{i_1 i_2 i_3 i_4 i_5} = \{z_{1,i_1}, z_{2,i_2}, z_{3,i_3}, z_{4,i_4}, z_{5,i_5}\}$$

$$c_{i_1 i_2 i_3 i_4 i_5} = -\ln p(Z_{i_1 i_2 i_3 i_4 i_5} \mid \tau_i)$$

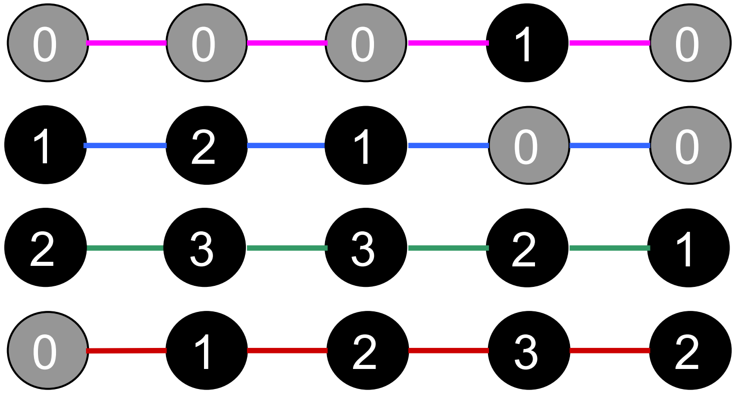
Data Association Problem



ω_i Partition hypothesis

$$C_{\omega_i} = -\ln p(Z \mid \omega_i)$$

$$= c_{00010} + c_{01232} + c_{12100} + c_{23321}$$



False alarm

Track termination

Track continuation

Track initiation

Multidimensional Assignment

$$c = \min \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \dots \sum_{i_T=0}^{n_T} c_{i_1 i_2 \dots i_T} x_{i_1 i_2 \dots i_T} \longrightarrow$$

Linear Cost Function

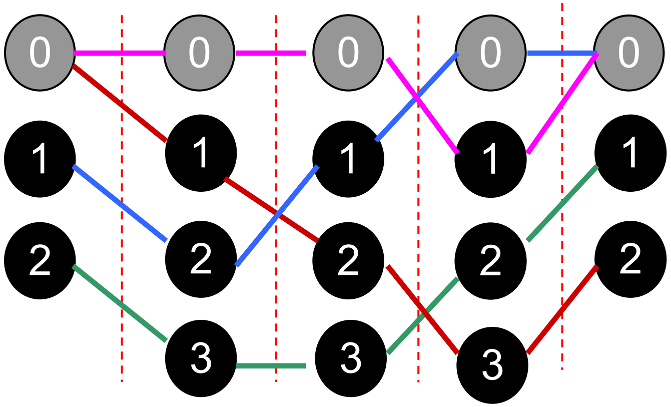
s.t.

$$\left.\begin{aligned}
&\sum_{i_2=0}^{n_2} \sum_{i_3=0}^{n_3} \dots \sum_{i_T=0}^{n_T} x_{i_1 i_2 \dots i_T} = 1; i_1 = 1, 2, \dots, n_1 \\
&\sum_{i_1=0}^{n_1} \sum_{i_3=0}^{n_3} \dots \sum_{i_T=0}^{n_T} x_{i_1 i_2 \dots i_T} = 1; i_2 = 1, 2, \dots, n_2 \\
&\qquad\vdots \\
&\sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \dots \sum_{i_{T-1}=0}^{n_{T-1}} x_{i_1 i_2 \dots i_T} = 1; i_N = 1, 2, \dots, n_T .
\end{aligned}\right\}$$

1. Exclusive and Exhaustive
2. No same origin

One-to-one correspondence

$$x_{i_1 i_2 \dots i_T} = \begin{cases} 1, & \text{if } z_{1,i_1}, z_{2,i_2} \dots, z_{T,i_T} \text{ forms a track} \\ 0, & \text{otherwise} \end{cases}$$



Two Technical Approaches for Multi-Obejct Multi-View Tracking

Tracking objects in 3D space with 2D measurements

- Reconstruction-Tracking Method

  First reconstruct 3D positions from multiple views, then apply tracking approach (feature-to-feature fusion)

  1. Find the correspondence across **views**
  2. Find the correspondence across **time**

Solution:

1. **Triangulation**
2. **3D Kalman filtering & 3D Data Association**

Two Technical Approaches for Multi-Obejct Multi-View Tracking

Tracking objects in 3D space with 2D measurements

- Reconstruction-Tracking Method

  First reconstruct 3D positions from multiple views, then apply tracking approach (feature-to-feature fusion)

  1. Find the correspondence across **views**
  2. Find the correspondence across **time**

- Tracking-Reconstruction Method

  First apply 2D tracking in each view independently, then reconstruct 3D trajectories through track-to-track associations (track-to-track fusion)

  1. Find the correspondence across **time**
  2. Find the correspondence across **views**

Two Technical Approaches for Multi-Obejct Multi-View Tracking

Tracking objects in 3D space with 2D measurements

- Tracking-Reconstruction Method

  First apply 2D tracking in each view independently, then reconstruct 3D trajectories through track-to-track associations (track-to-track fusion)

  1. Find the correspondence across **time**
  2. Find the correspondence across **views**

Solution by Wu et al., ICCV 2009

1. **Iterative GRASP**
2. **Information Fusion**

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking

– A Multidimensional Assignment Formulation (1)

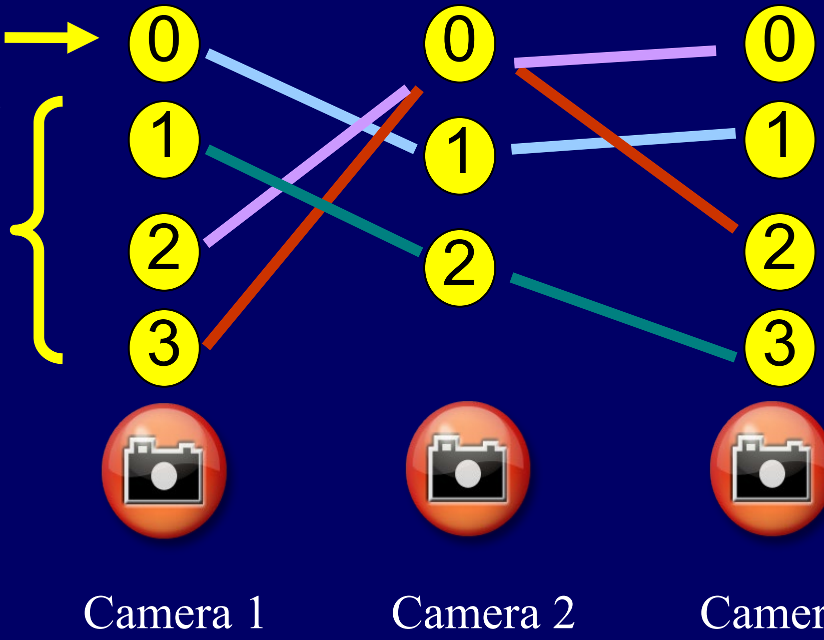
Given N calibrated and synchronized cameras that share overlapping fields of view and n_s measurements in the field of view of camera s, the state $x^{(t)}$ (3D coordinates) of an object of interest at time t can be assumed to evolve in time according to the equations:

$$x^{(t+1)} = Ax^{(t)} + v^{(t)}$$

$$z_{s,i_s}^{(t)} = H_s \, x^{(t)} + w^{(t_s)} \; ; s = 1, \ldots, N; \; i_s = 1, \ldots, n_s$$

where $v^{(t)}$ and $w^{(t_s)}$ are independent zero-mean Gaussian noise processes with respective covariances Q(t) and $R_s(t)$, A is the state transition matrix, and H_s the projection matrix for camera s.

Multi-object Multi-view Tracking
– A Multidimensional Assignment Formulation (2)



Dummy measurement to represent missing detection

Real Measurements

Camera 1

Camera 2

Camera 3

Cost for this 3-tuple:

$$c_{z_{1,1}z_{2,2}z_{3,3}} = -\ln \frac{p(z_{1,1}z_{2,2}z_{3,3}\,|\,a)}{p(z_{1,1}z_{2,2}z_{3,3}\,|\,\%)}$$

← Likelihood the measurements describe some object

← Likelihood the measurements are all false positives

%=dummy object

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking

– A Multidimensional Assignment Formulation (3)

The likelihood that measurements $z_{1,i1}, z_{2,i2}, \ldots, z_{N,iN}$ ($=Z_{i1i2\ldots iN}$) describe object state x_a is given as

$$p(Z_{i_1 i_2 \ldots i_N} \mid x_a) = \prod_{s=1}^{N} \{[1 - \mathrm{P}_{D_s}]^{1-u(i_s)} \times [\mathrm{P}_{D_s}\, p(z_{s,i_s} \mid x_a)]^{u(i_s)}\}$$

$$u(i_s) = \begin{cases} 0 & \text{if } \mathrm{i}_\mathrm{s} = 0, \\ 1 & \text{otherwise} \end{cases}$$

$$p(z_{s,i_s} \mid x_a) = \mathsf{N}\,(z_{s,i_s}; H_s\, x_a, R_s)$$

where P_{Ds} is the detection rate; $z_{s,0}$ means object is not detected in camera s

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking
– A Multidimensional Assignment Formulation (4)

The likelihood that measurements $z_{1,i1}$, $z_{2,i2}$, …, $z_{N,iN}$ ($Z_{i1i2\ldots iN}$) are unrelated to any object is given as

$$p(Z_{i_1 i_2 \ldots i_N} \mid \%) = \prod_{s=1}^{N} [\frac{1}{\Phi_s}]^{u(i_s)}$$

where Φs is the volume of FOV in camera s

We now can define the cost of associating N-tuple $Z_{i1i2\ldots iN}$ to object a at time t is as the negative log-likelihood ratio:

$$c_{i_1 i_2 \ldots i_N} = -\ln \frac{p(Z_{i_1 i_2 \ldots i_N} \mid a)}{p(Z_{i_1 i_2 \ldots i_N} \mid \%)}$$

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking

– A Multidimensional Assignment Formulation (5)

We do not know the true state x_a in the likelihood Equation

$$p(z_{s,i_s} \mid x_a) = \mathrm{N}\,(z_{s,i_s}; H_s\, x_a, R_s)$$

so we replace it by $\hat{x}_a = \arg\min_{x_a} \sum_{s=1}^{n} d(z_{s,i_s}, H_s\, x_a)$

where d is Euclidean distance between $H_s x_a$, the object position projected onto the image s, and the corresponding measurement $z_{s,is}$

This is a 3D reconstruction problem.

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking

– A Multidimensional Assignment Formulation (6)

Assuming that such associations are independent, our goal is to find the most likely set of N-tuples that minimizes the linear cost function

$$
\begin{aligned}
&c = \min \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \cdots \sum_{i_N=0}^{n_N} c_{i_1 i_2 \ldots i_N} x_{i_1 i_2 \ldots i_N} \\
&\text{s.t.} \\
&\sum_{i_2=0}^{n_2} \sum_{i_3=0}^{n_3} \cdots \sum_{i_N=0}^{n_N} x_{i_1 i_2 \ldots i_N} = 1; i_1 = 1, 2, \ldots, n_1 \\
&\sum_{i_1=0}^{n_1} \sum_{i_3=0}^{n_3} \cdots \sum_{i_N=0}^{n_N} x_{i_1 i_2 \ldots i_N} = 1; i_2 = 1, 2, \ldots, n_2 \\
&\qquad \vdots \\
&\sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \cdots \sum_{i_{N-1}=0}^{n_{N-1}} x_{i_1 i_2 \ldots i_N} = 1; i_N = 1, 2, \ldots, n_N .
\end{aligned}
$$

NP Hard

where we use binary variable $x_{i1i2\ldots iN}$ to indicate if $Z_{i1i2\ldots iN}$ is associated with a candidate object or not.

Tracking Reconstruction Method by Wu et al., ICCV 2009

Multi-object Multi-view Tracking

Greedy Randomized Adaptive Search Procedure (GRASP)

Multistart random process [Feo 1989, Renata 2003]

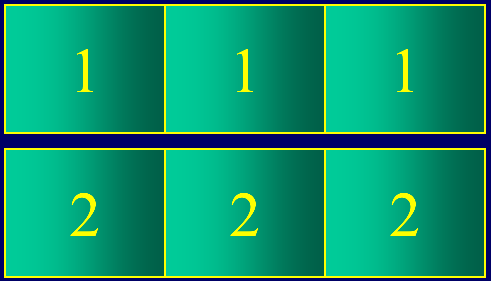
Within each iteration,

- Construction Phase

  Randomly construct a feasible greedy solution

- Local Search Phase

  Improve the feasible solution by local search in the neighborhood

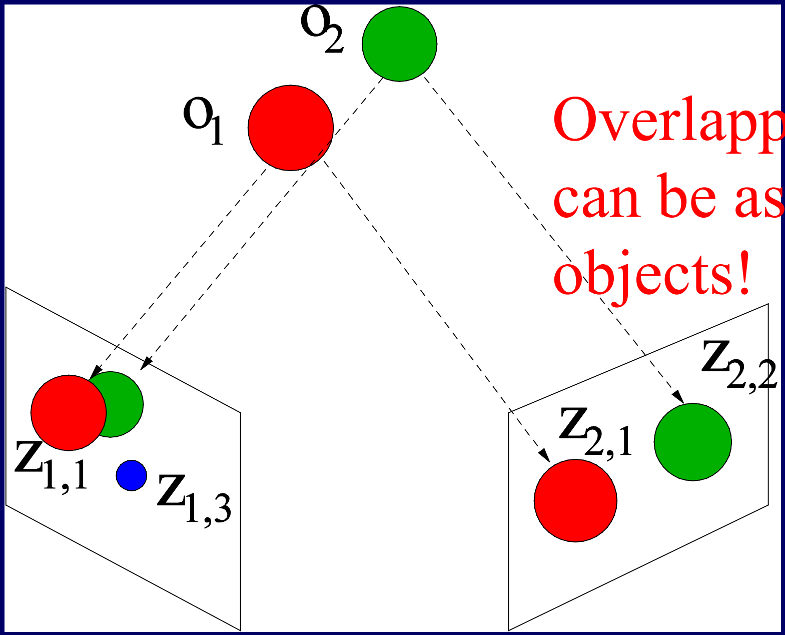


1	1	1
2	2	2

Multi-object Multi-view Tracking

— Iterative Greedy Randomized Adaptive Search Procedure (1)

The constraints in the multidimensional assignment formulation imply the one-to-one correspondence between measurements and objects, which is not desirable in the multi-view tracking scenario because of occlusion.



Overlapping measurements can be assigned to multiple objects!

Multi-object Multi-view Tracking

— Iterative Greedy Randomized Adaptive Search Procedure (2)

Define all possible N-tuples as $F = Z_1 \times Z_2 \times \ldots Z_N$, Z_i which is the set of all measurements in camera s plus a "dummy" measurement that represents a missing detection event.

We divide the set of the suboptimal assignments Z found by GRASP into two subsets:

- Confirmed associations:

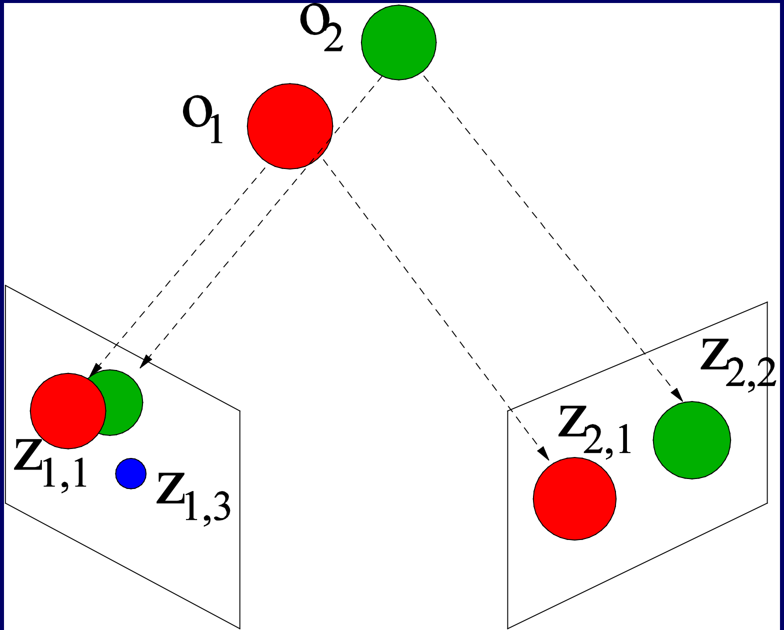
$$M_c = \{Z_{i_1 i_2 \ldots i_N} \mid x_{i_1 i_2 \ldots i_N} = 1; i_1 \neq 0; \ldots; i_N \neq 0\}$$

- Suspicious associations (dummy measurements & objects):

$$M_s = Z \setminus M_c$$

Multi-object Multi-view Tracking

– Iterative Greedy Randomized Adaptive Search Procedure (3)



Iteration 1:

$F = \{(z_{1,1}\ z_{2,0}), (z_{1,1}\ z_{2,1}), (z_{1,1}\ z_{2,2}), (z_{1,3}\ z_{2,0}), (z_{1,3}\ z_{2,1}), (z_{1,3}\ z_{2,2}), (z_{1,0}\ z_{2,1}), (z_{1,0}\ z_{2,2})\}$

$Z = \{(z_{1,1}\ z_{2,1}), (z_{1,3}\ z_{2,0}), (z_{1,0}\ z_{2,2})\}$

$M_c = \{(z_{1,1}\ z_{2,1})\}$, $M_s = \{(z_{1,3}\ z_{2,0}), (z_{1,0}\ z_{2,2})\}$

Iteration 2:

$F = \{(z_{1,1}\ z_{2,0}), (z_{1,1}\ z_{2,2}), (z_{1,3}\ z_{2,0}), (z_{1,3}\ z_{2,1}), (z_{1,3}\ z_{2,2}), (z_{1,0}\ z_{2,1}), (z_{1,0}\ z_{2,2})\}$

$Z = \{(z_{1,1}\ z_{2,2}), (z_{1,3}\ z_{2,0}), (z_{1,0}\ z_{2,1})\}$

$M_c = \{(z_{1,1}\ z_{2,2})\}$, $M_s = \{(z_{1,3}\ z_{2,0}), (z_{1,0}\ z_{2,1})\}$

Multi-object Multi-view Tracking

— Iterative Greedy Randomized Adaptive Search Procedure (4)

IGRASP

Building Phase

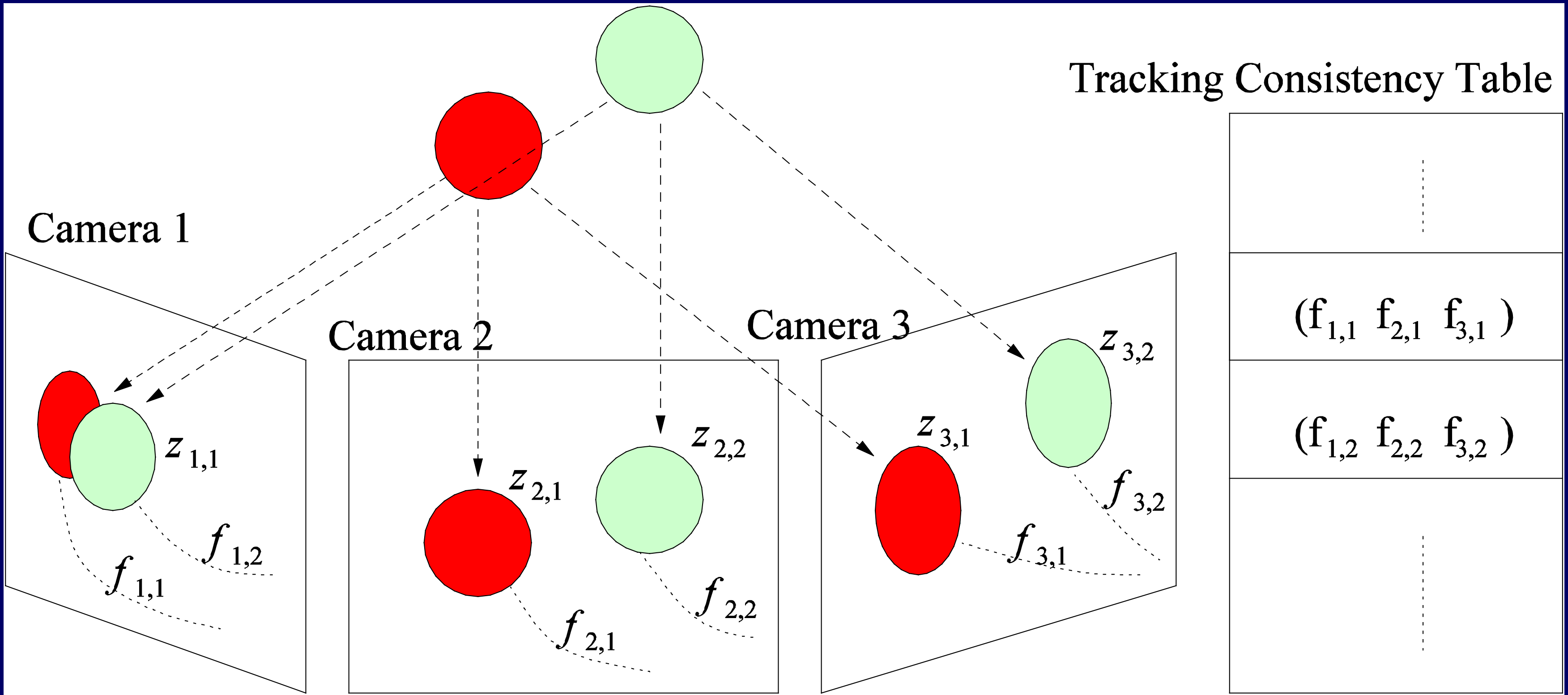
Initialization by computing the costs for all possible associations in set F

Solving Phase

```
For i = 1, …, maxiter,
1.    Formulate multidimensional assignment problem on set F
2.    Solve the problem by GRASP
3.    Partition the computed solution into confirmed set Mc and suspicious set Ms
4.    If set Mc is empty, terminate;
      Else F = F \ Mc
End
Output the final suboptimal solution.
```

Multi-object Multi-view Tracking

– Fusion of Information from Multiple Views (1)



Tracking Consistency Table

⋮
($f_{1,1}$ $f_{2,1}$ $f_{3,1}$)
($f_{1,2}$ $f_{2,2}$ $f_{3,2}$)
⋮

$f_{i,j}$= j^{th} 2D Kalman tracker in view i; $z_{i,j}$= j^{th} measurement in view i

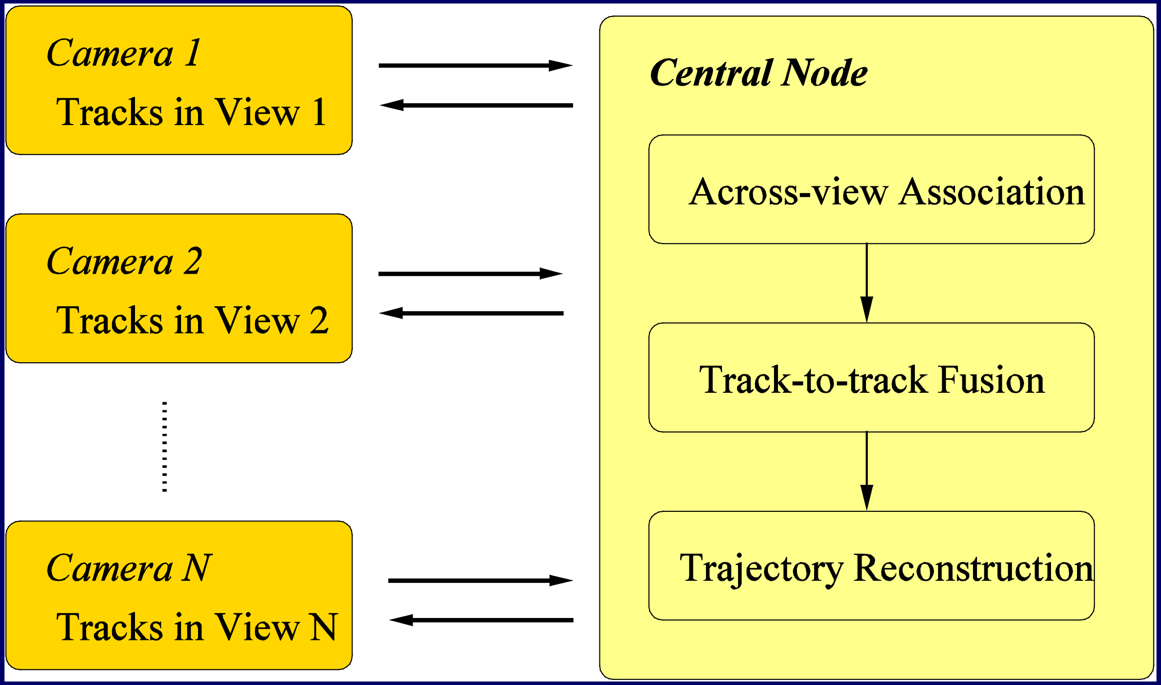
Multi-object Multi-view Tracking

— Fusion of Information from Multiple Views (2)

Tracking is performed at each camera.

Tracks and measurements are sent to a central node for processing.

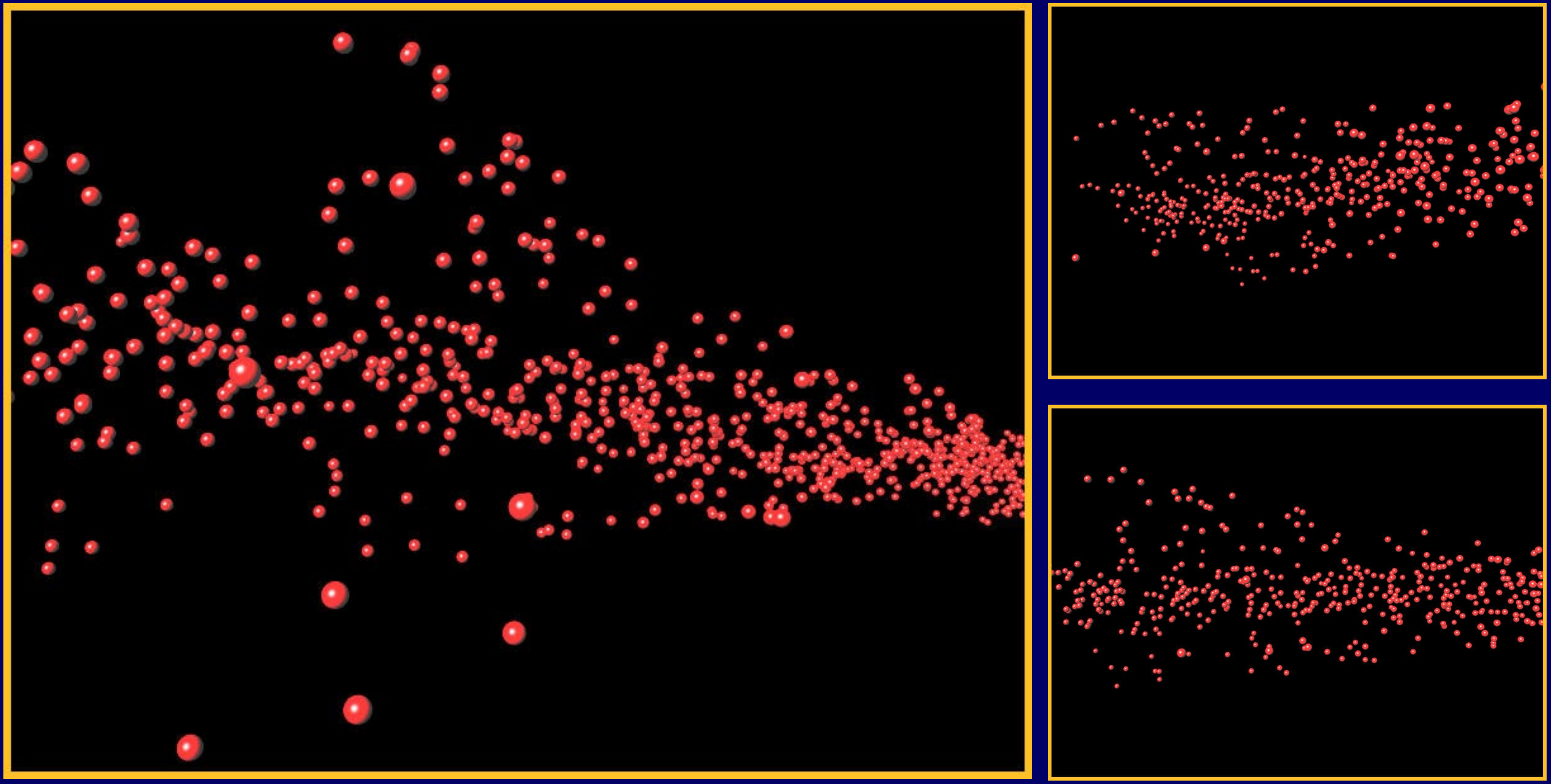
Each camera tracker adjusts its across-time associations based on the fusion result it receives from the central node



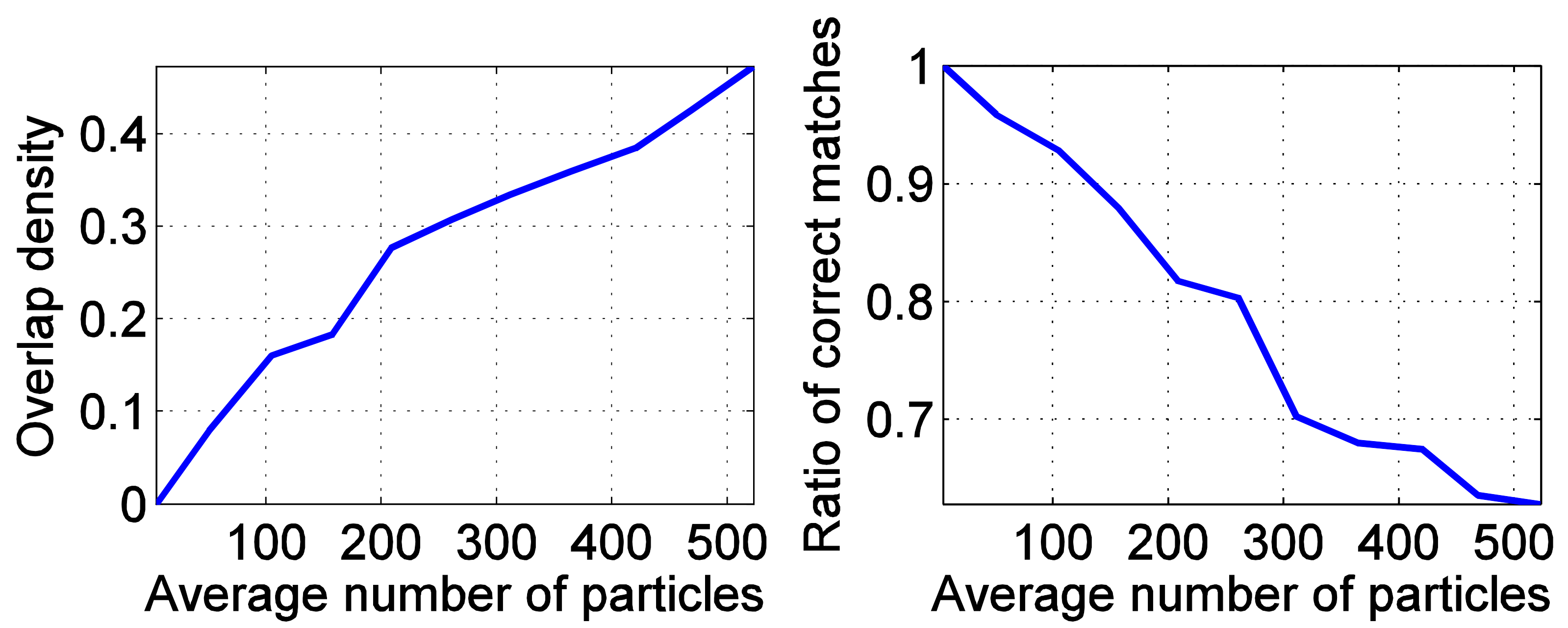
Experiment – Validation of across-view association (1)

Synthetic Data:

- randomly generate spherical particles of radius 28cm to move in 20×5×5m^3 space at a fixed speed of 2m/s
- 10 datasets with increasing emergence rates between 1 and 100 particles/sec



Experiment – Validation of across-view association (2)

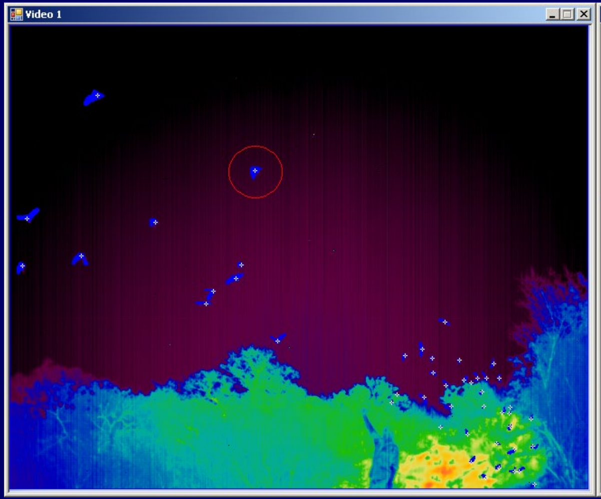


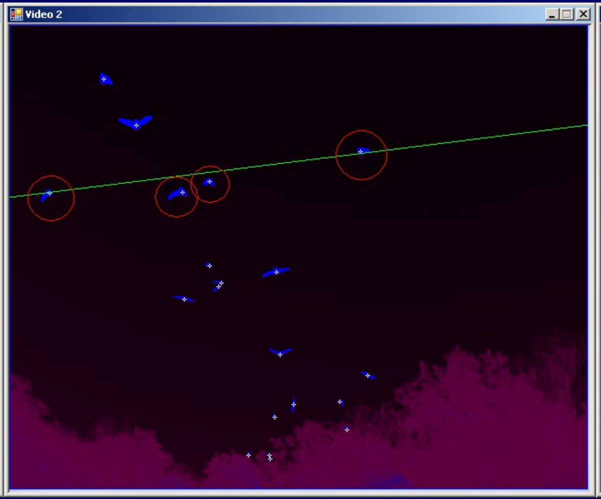
Overlap density: the ratio of number of overlapping particle projections over the total number of particles

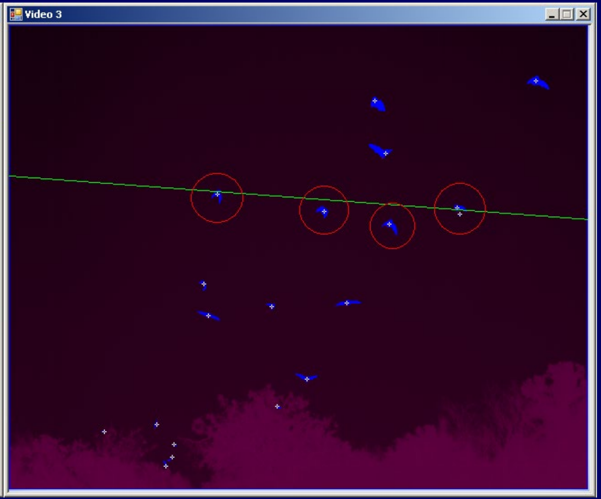
Ratio of correct matches: number of correct tuples found by IGRASP over the ground truth

Experiment – Validation of across-view association (3)

- Make the problem as sparse as possible!!! Evaluate the candidate tuples that lie within the neighborhood of corresponding epipolar lines

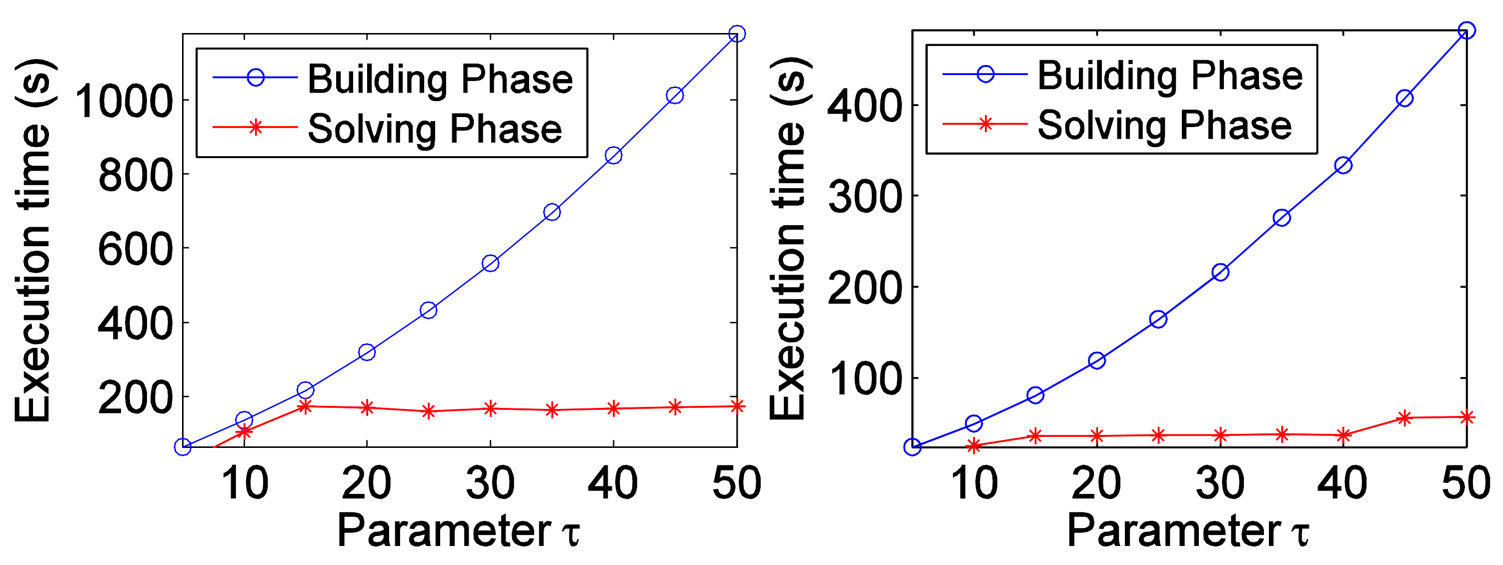




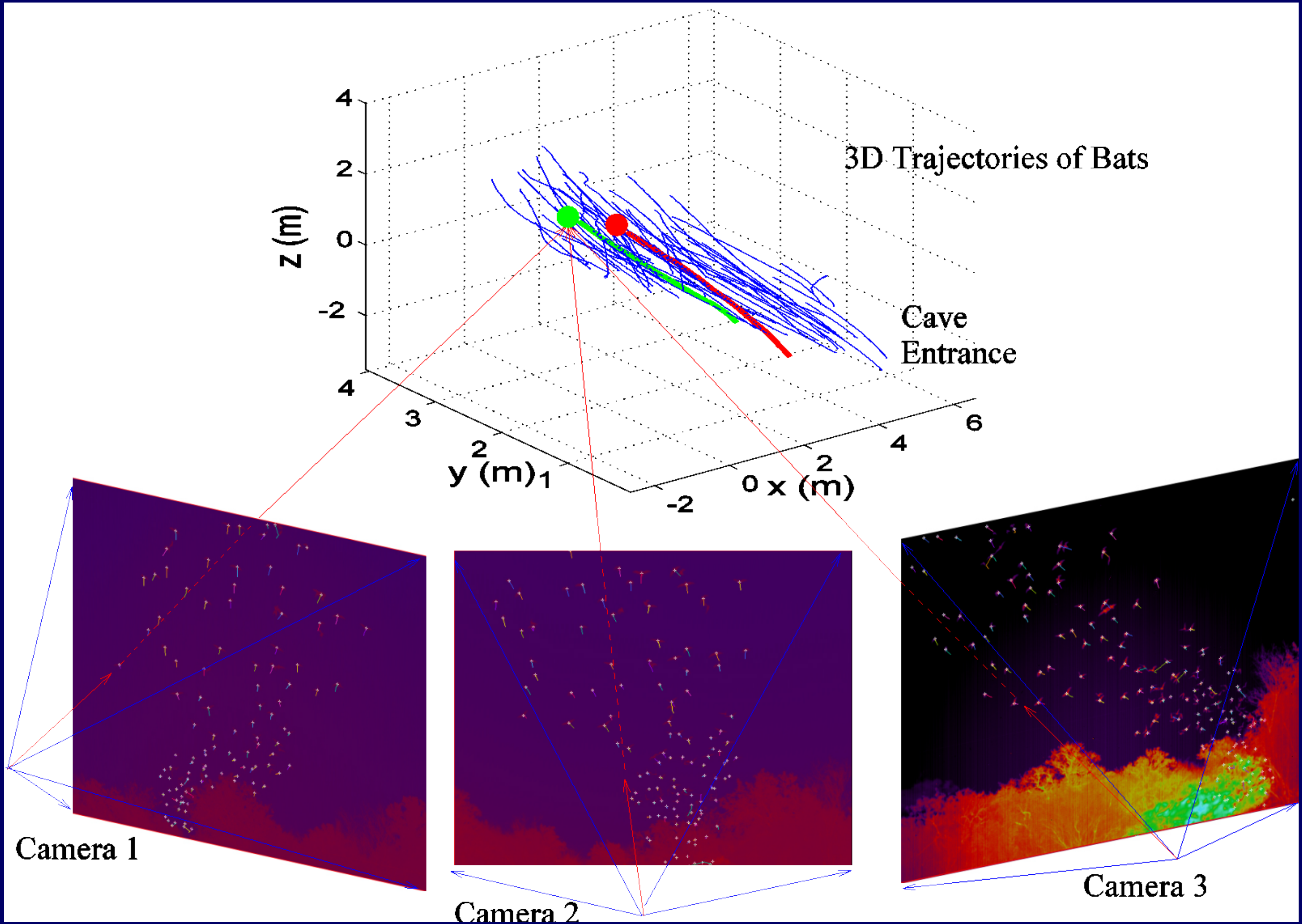


Experiment – Validation of across-view association (4)

- Make the problem as sparse as possible!!! Evaluate the candidate tuples that lie within the neighborhood of corresponding epipolar lines



Experiment – Infrared Thermal Video Analysis (1)



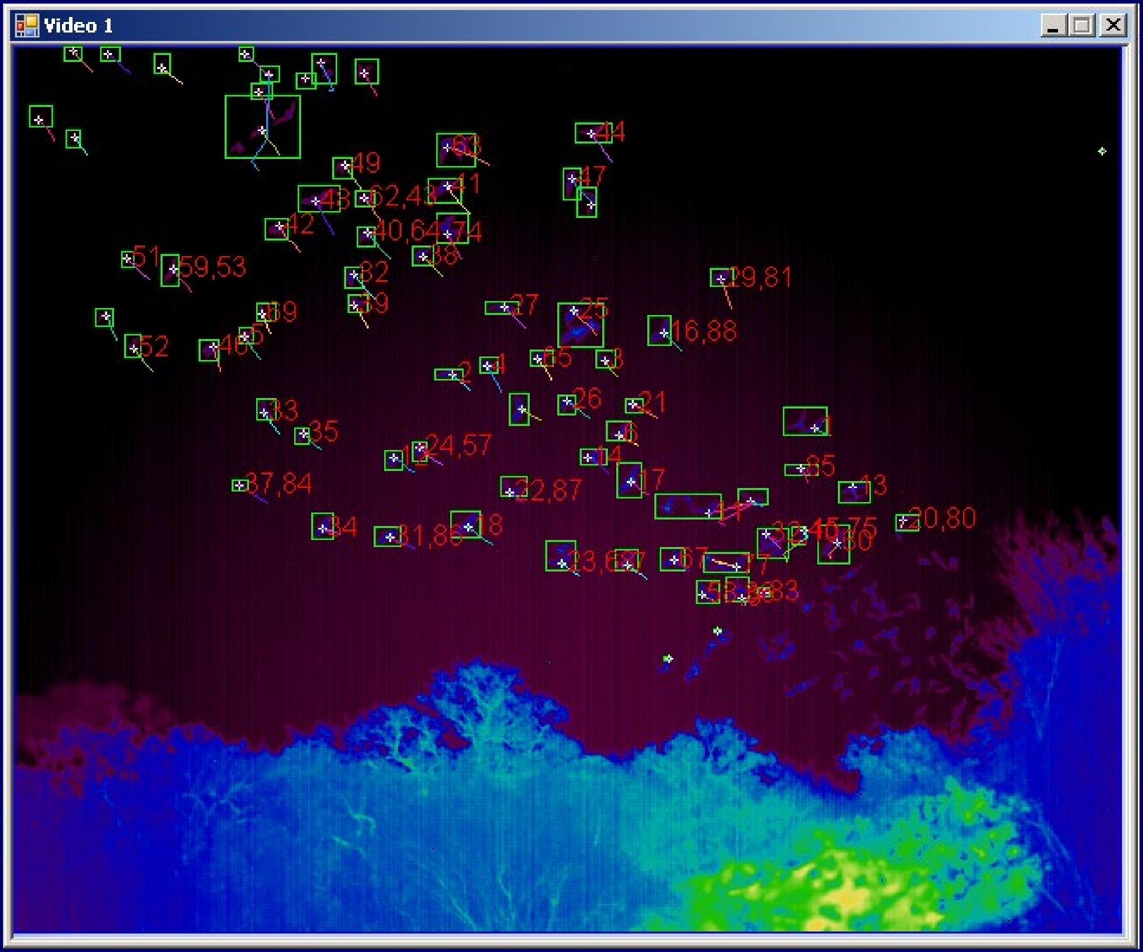
Experiment – Infrared Thermal Video Analysis (2)

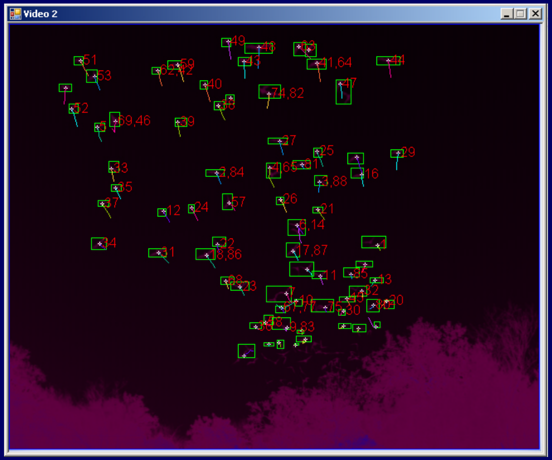
- 4 test sequences with different density levels of the column of emerging bats; each sequence has 100 frames.

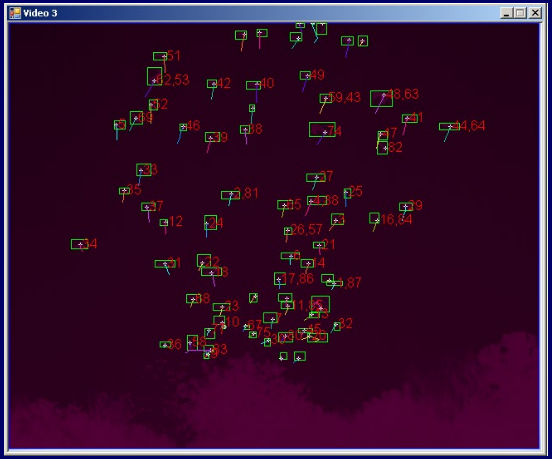
Number of Bat/frame	True Number of Bats	Computed Number of Tracks	Number of Occlusion	Number of Recovered Occlusions
20	25	33	56	40
40	50	63	94	54
60	71	90	140	86
100	119	185	368	88

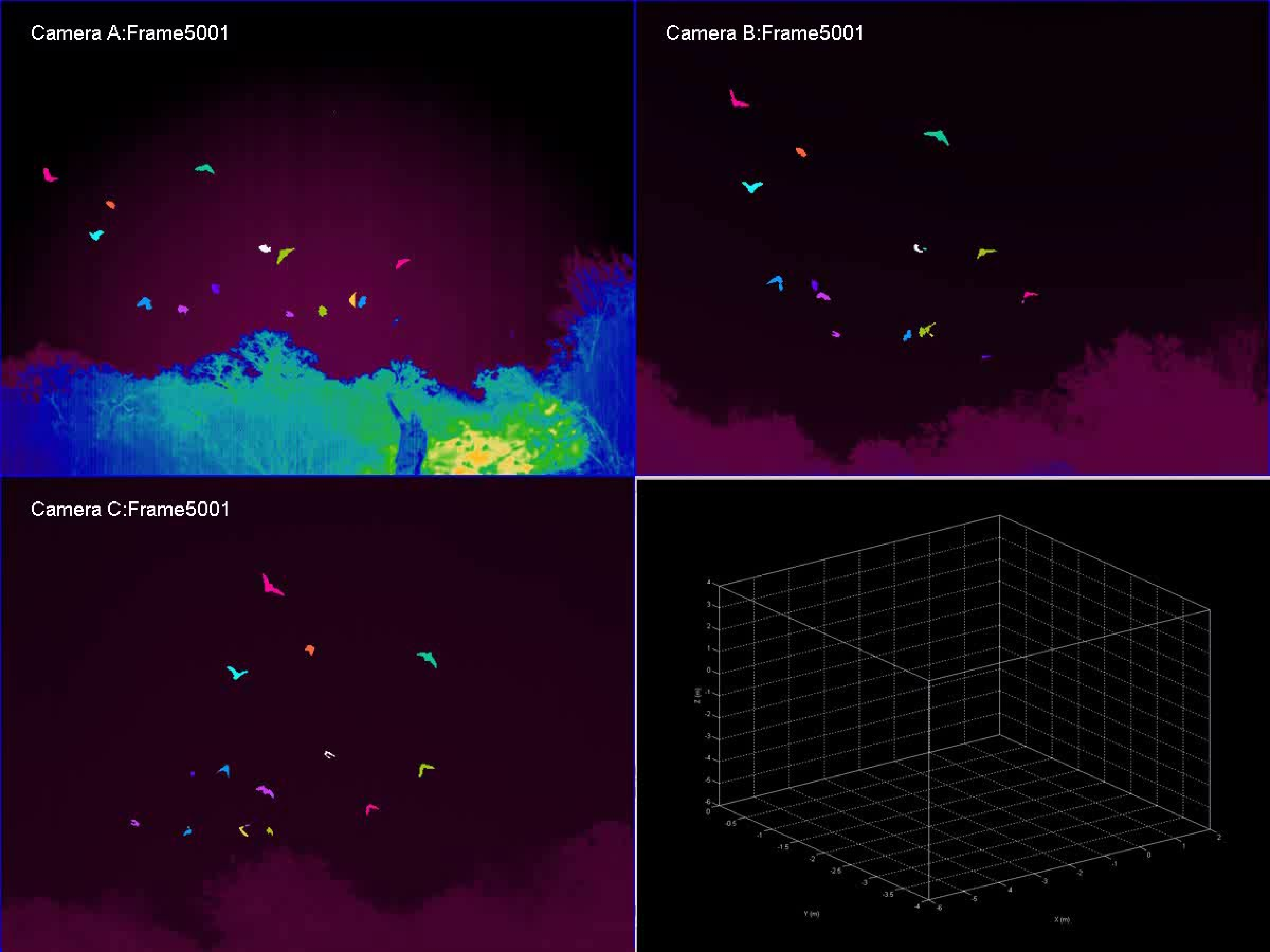
24% correctly interpretation …
Occlusions happen in more than one view

Experiment – Infrared Thermal Video Analysis (3)

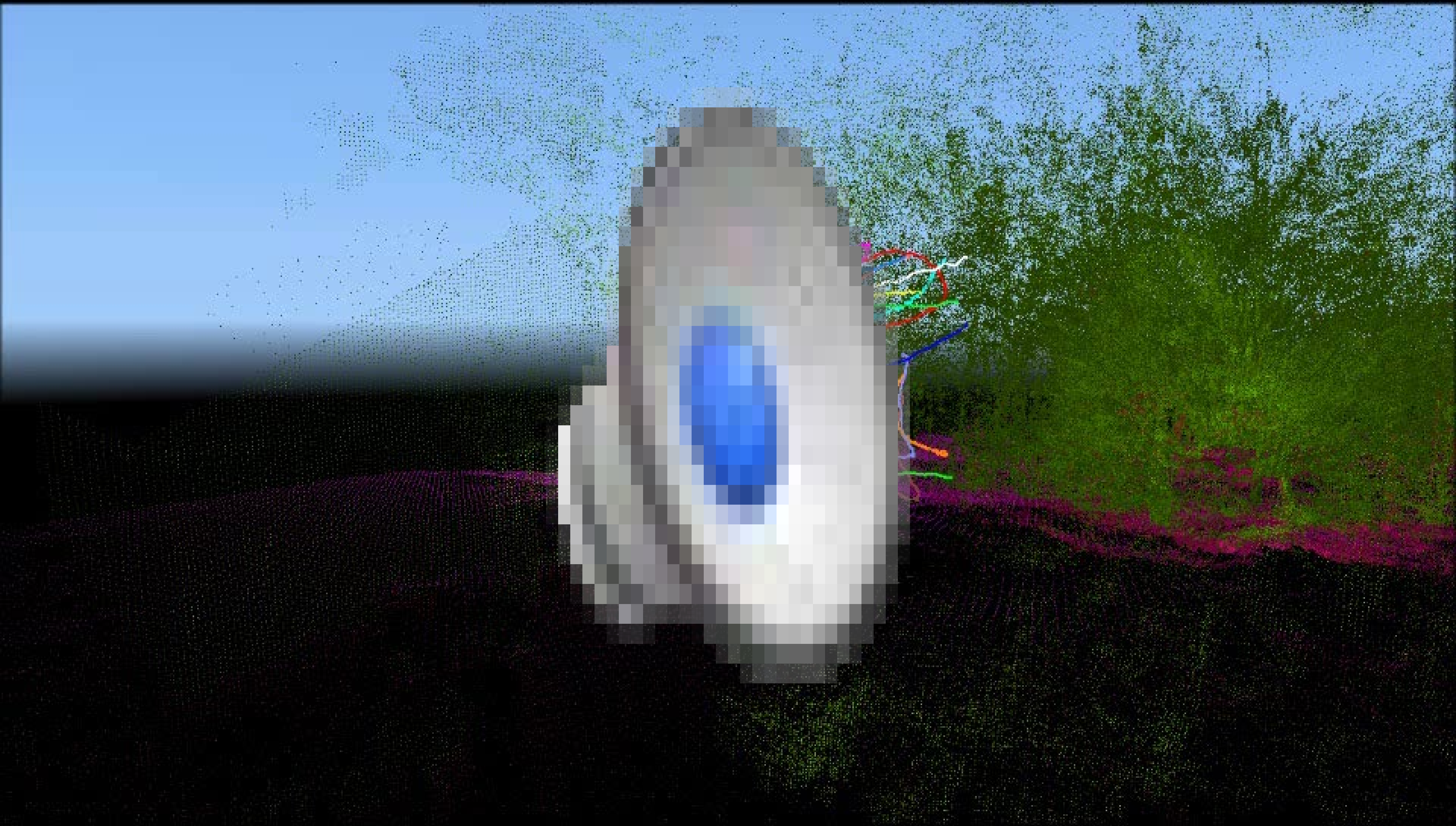







Camera A:Frame5001
Camera B:Frame5001
Camera C:Frame5001
4
3
2
1
0
-1
-2
-3
-4
-5
-6
Z (m)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
Y (m)
-6
-5
-4
-3
-2
-1
0
1
2
X (m)

Multi modal reconstruction



Xiaoyuan Yang
Boston University/UMass Boston

MultiView MultiObject Tracking State of the Art in 2024

- Deep learning used for detection of objects
- Measurement Association/Track Association still done by traditional algorithms

- Survey (incomplete): Amosa et al., 2023
- Tracking-reconstruction & reconstruction-tracking paradigms applied to people tracking: Yang et al., 2022

Learning Outcomes: Be able to

- Define the Multi-view Multi-object Tracking problem
- Explain Difference between Tracking-Reconstruction and Reconstruction-Tracking Methods and possible solutions
- Discuss how multi-view multi-object tracking systems can be validated experimentally