BOSTON UNIVERSITY

Tracking a Large Number of Objects from Multiple Views

Zheng Wu[1], Nickolay. Hristov[1], Tyson L. Hedrick[2], Thomas H. Kunz[1], Margrit Betke[1]
[1]Boston University, [2]University of North Carolina at Chapel Hill, USA

Problem

Tracking large numbers of tightly-spaced objects that rapidly move in three dimensions with multiple views.

Contribution

- A multidimensional assignment formulation for across-view data-association in large crowds based on multi-view geometry.
- An iterative procedure to solve the across-view association problem.
- A stereoscopy method to reconstruct the trajectories of objects moving in 3D space that employs epipolar-neighborhood search.
- A new information fusion technique that ensures interpretation of occlusion and consistency of tracking.

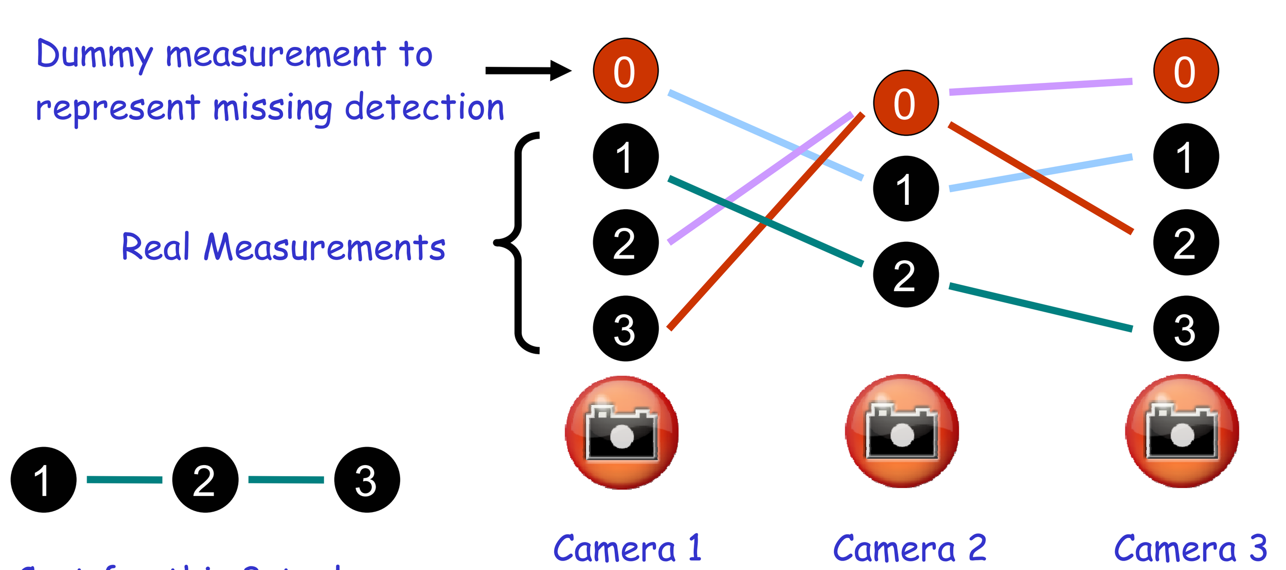
Multi-object Multi-view Tracking

• A Multidimensional Assignment Formulation

Given N calibrated and synchronized cameras 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:

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

where $v^{(t)}$ and $w^{(ts)}$ 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.



Cost for this 3-tuple:

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

← Likelihood the measurements describe some object
← Likelihood the measurements are all false positives

Use binary variable $x_{i1i2\ldots iN}$ to indicate if $(z_{1,i1}, z_{2,i2}, \ldots, z_{N,iN})$ is associated with an object or not. The goal is to minimize the linear function:

$$c = \min \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \ldots \sum_{i_N=0}^{n_N} c_{i_1 i_2 \ldots i_N} x_{i_1 i_2 \ldots i_N} \qquad (2)$$

s.t.

$$\sum_{i_2=0}^{n_2} \sum_{i_3=0}^{n_3} \ldots \sum_{i_N=0}^{n_N} x_{i_1 i_2 \ldots i_N} = 1;\ i_1 = 1,2,\ldots,n_1$$
$$\sum_{i_1=0}^{n_1} \sum_{i_3=0}^{n_3} \ldots \sum_{i_N=0}^{n_N} x_{i_1 i_2 \ldots i_N} = 1;\ i_2 = 1,2,\ldots,n_2$$
$$\vdots$$
$$\sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \ldots \sum_{i_{N-1}=0}^{n_{N-1}} x_{i_1 i_2 \ldots i_N} = 1;\ i_N = 1,2,\ldots,n_N$$

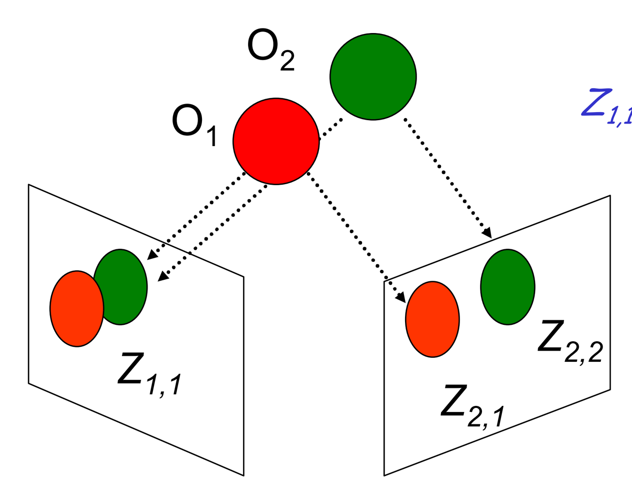
One-to-one correspondence: Each measurement has to be assigned and assigned only once!

• Greedy Randomized Adaptive Search Procedure (GRASP)

A multistart random process [Feo and Resende 1995], within each iteration,

- *Construction Phase* : Randomly construct a feasible greedy solution
- *Local Search Phase* : Improve the feasible solution by local search

• Iterative Greedy Randomized Adaptive Search Procedure



$Z_{1,1}$ has to be assigned twice!!!

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

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

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

Building Phase **:**

Define all tuples as $F = Z_1 \times Z_2 \times \ldots Z_N$, Z_i is the set of measurements in view s.
Initialization by computing the costs for all possible associations in set F.

Solving Phase **:**

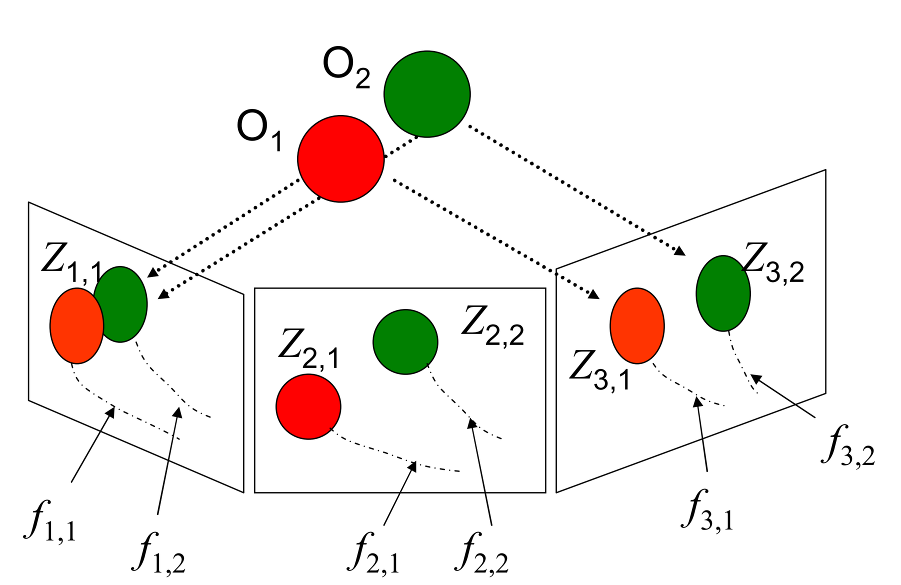
For i = 1, …, *maxiter*,

- Formulate multidimensional assignment problem on set F
- Solve the problem by GRASP to get a suboptimal solution Z
- We divide the set of the assignments Z into two subsets:
  - *Confirmed associations*: $M_c = \{Z_{i_1 i_2 \ldots i_N} \,|\, x_{i_1 i_2 \ldots i_N} = 1; i_1 \neq 0; \ldots; i_N \neq 0\}$
  - *Suspicious associations*: $M_s = Z \setminus M_c$
- Partition the computed solution into set Mc and Ms
- If set Mc is empty, terminate; Else F = F \ Mc

End

Output the final suboptimal solution.

• Fusion of Information from Multiple Views



Track-to-track association table

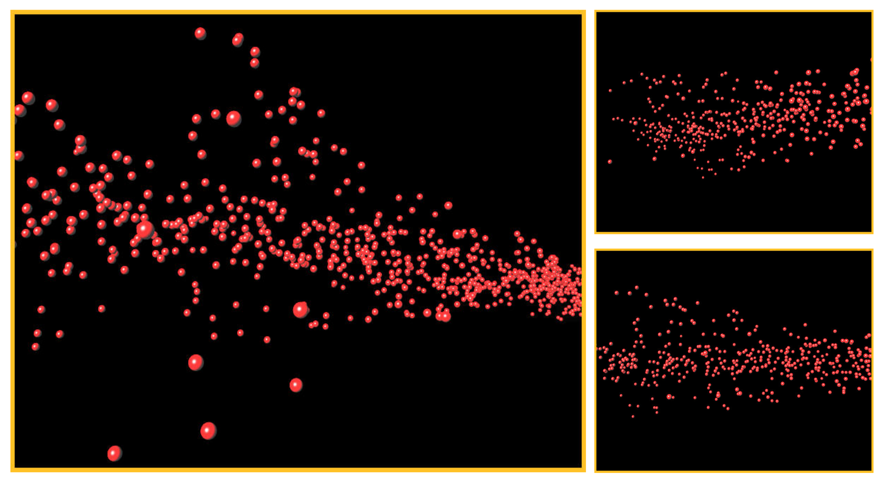
$(f_{1,1}\ f_{2,1}\ f_{3,1})$
$(f_{1,2}\ f_{2,2}\ f_{3,2})$
⋮

$f_{i,j}$: jth tracker in view i $\quad z_{i,j}$: jth measurement in view i

Tracking is performed at each sensor level and tracks and measurements are sent to a central node for processing. Each sensor tracker adjusts its across-time associations based on the fusion result it receives from the central node. i.e. tracker that claims lost in some view can be recovered by checking the states of associated trackers in other views.

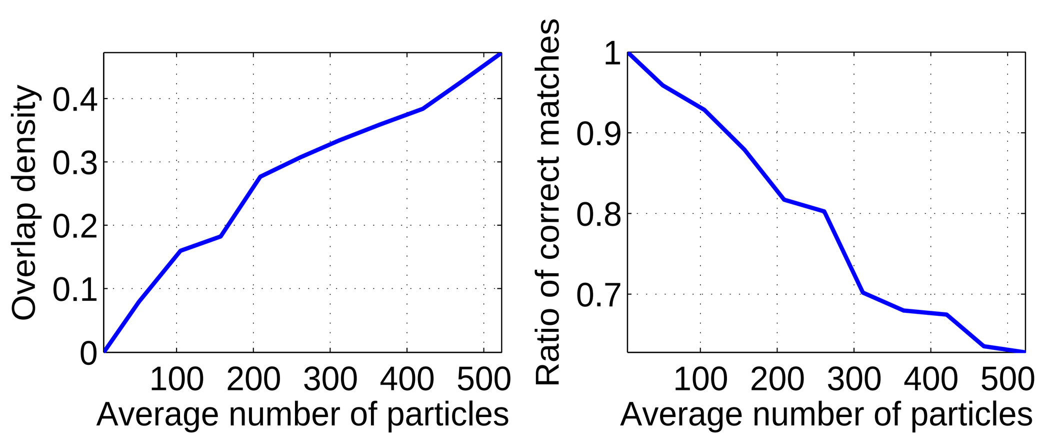
Experiment 1: Synthetic Data

- randomly generate spherical particles of radius 28cm to move in 20×5×5m³ space at a fixed speed of 2m/s. The 3D coordinates of each particle is known
- 10 datasets with increasing emergence rates between 1 and 100 particles/sec

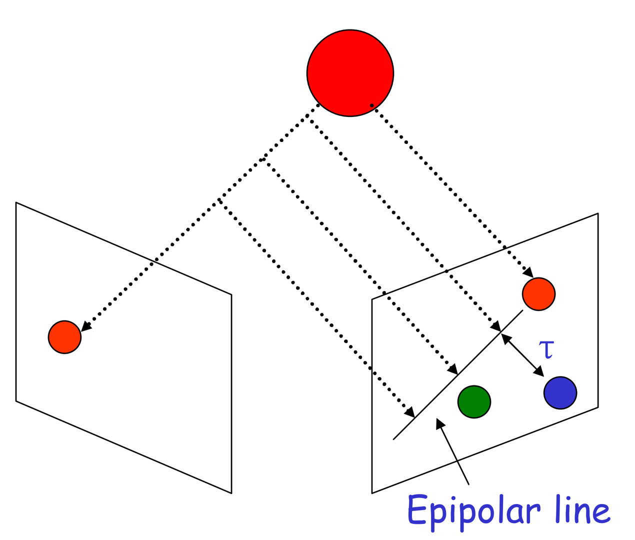


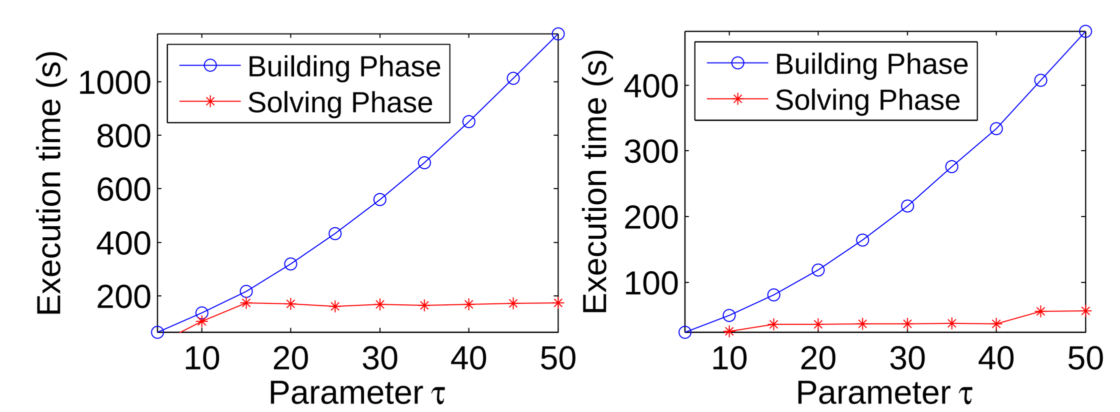
Overlap density: the ratio of number of overlapping particle projections over the total number of particles

Ratio of correct matches: number of correct across-view associations found by IGRASP over the ground truth



The bottleneck of computation is to compute the costs of all possible tuples. Make the problem as sparse as possible!!! Evaluate the candidate tuples that lie within the neighborhood of corresponding epipolar lines. The size of the neighborhood is determined by the perpendicular distance (τ) between measurement and epipolar line.





Execution time of IGRASP (Matlab version) with different values of τ for the across-view association. Left: 100 particles/s. Right: 50 particles/s.

Experiment 2: Infrared Thermal Video Analysis

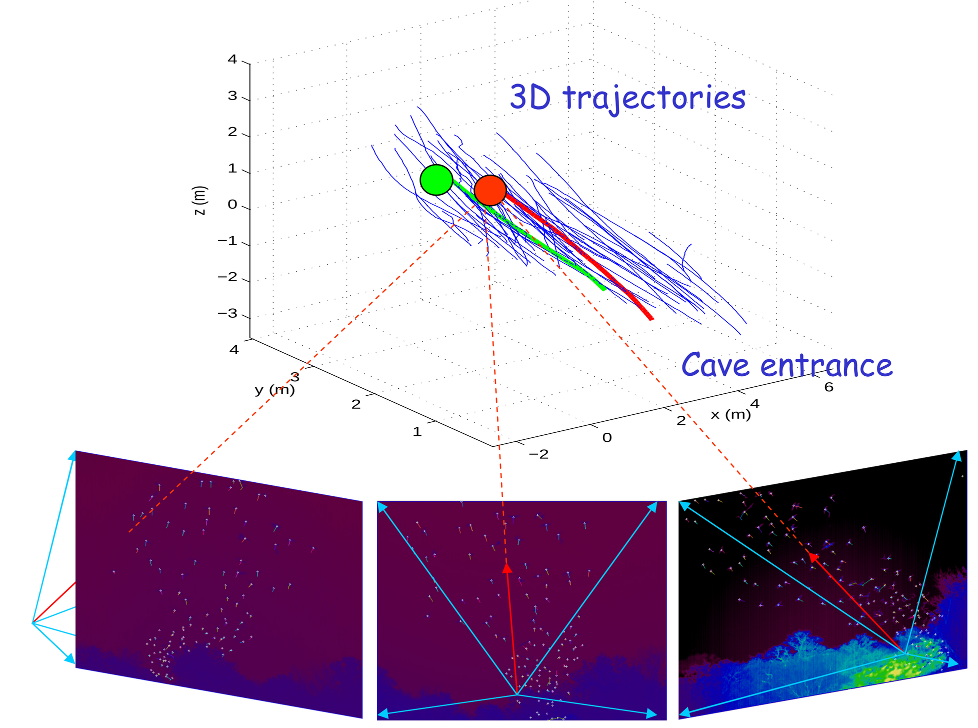


Table 1. Performance of tracking system in resolving occlusions. Ground truth was established by manual marking of four 100-frame sequences.

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

24% correctly interpretation, because occlusions happen in more than one view

Reference

Feo, T. A., M. G. C. Resende. Greedy randomized adaptive search procedures. *J. Global Optim.* 6, 109-133, 1995