Optimal Multi-View Fusion of Object Locations

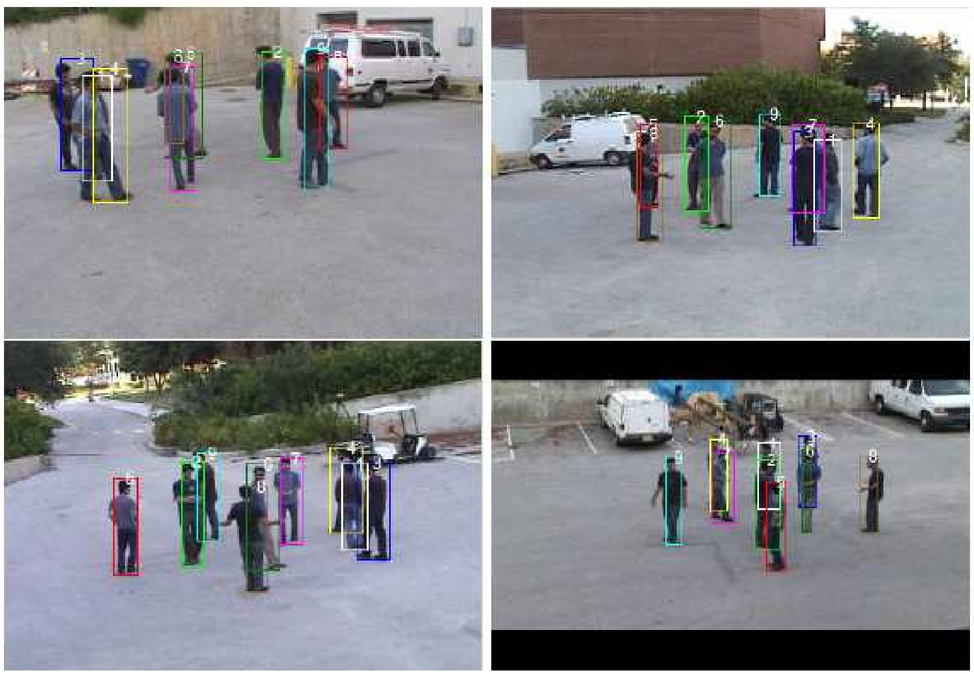
Aswin C. Sankaranarayanan
Rama Chellappa

Center for Automation Research
University of Maryland, College Park

UMIACS
UNIVERSITY OF MARYLAND INSTITUTE FOR ADVANCED COMPUTER STUDIES

Problem Setting

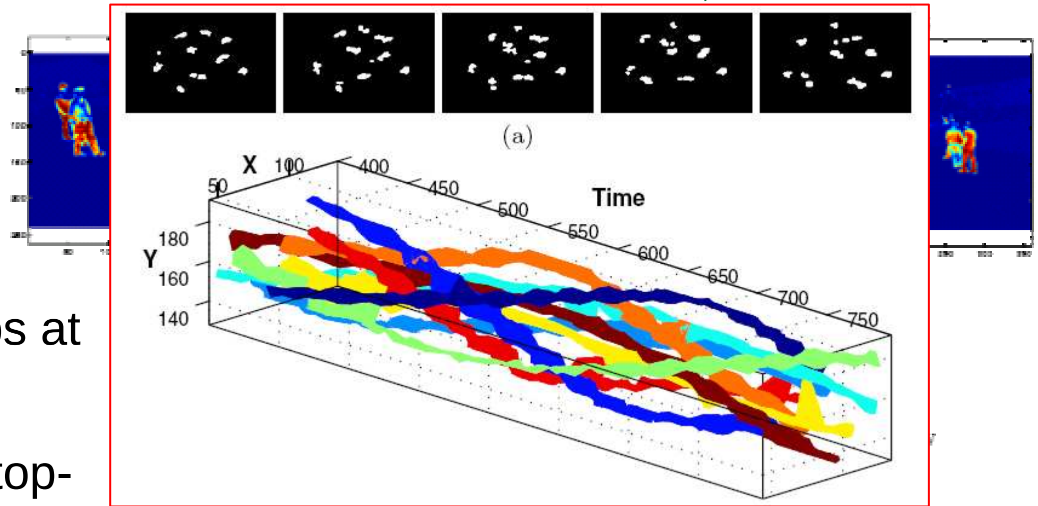
- Multiple cameras observing a plane.
  - Tracking on the plane.



[Khan, Shah, ECCV 2006]

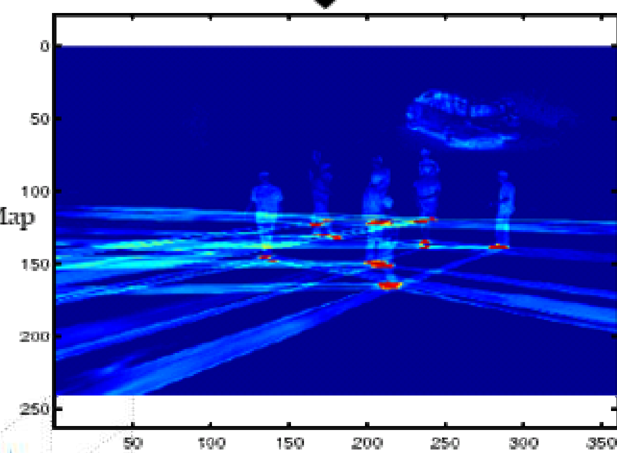
x-y track on the ground plane information can be useful.

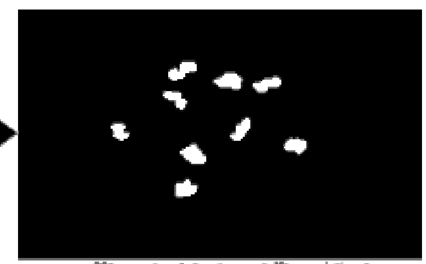
Prior Work (Khan and Shah, ECCV 2006)



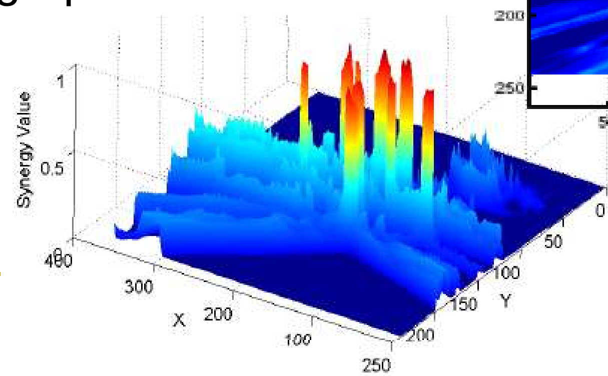
- Foreground likelihood maps at each view.
- Project map to reference (top-down) view.
- Combine to obtain a Synergy map.
- Threshold to obtain leg locations.
- Stack such images temporally and segment using graph-cuts.

Synergry Map



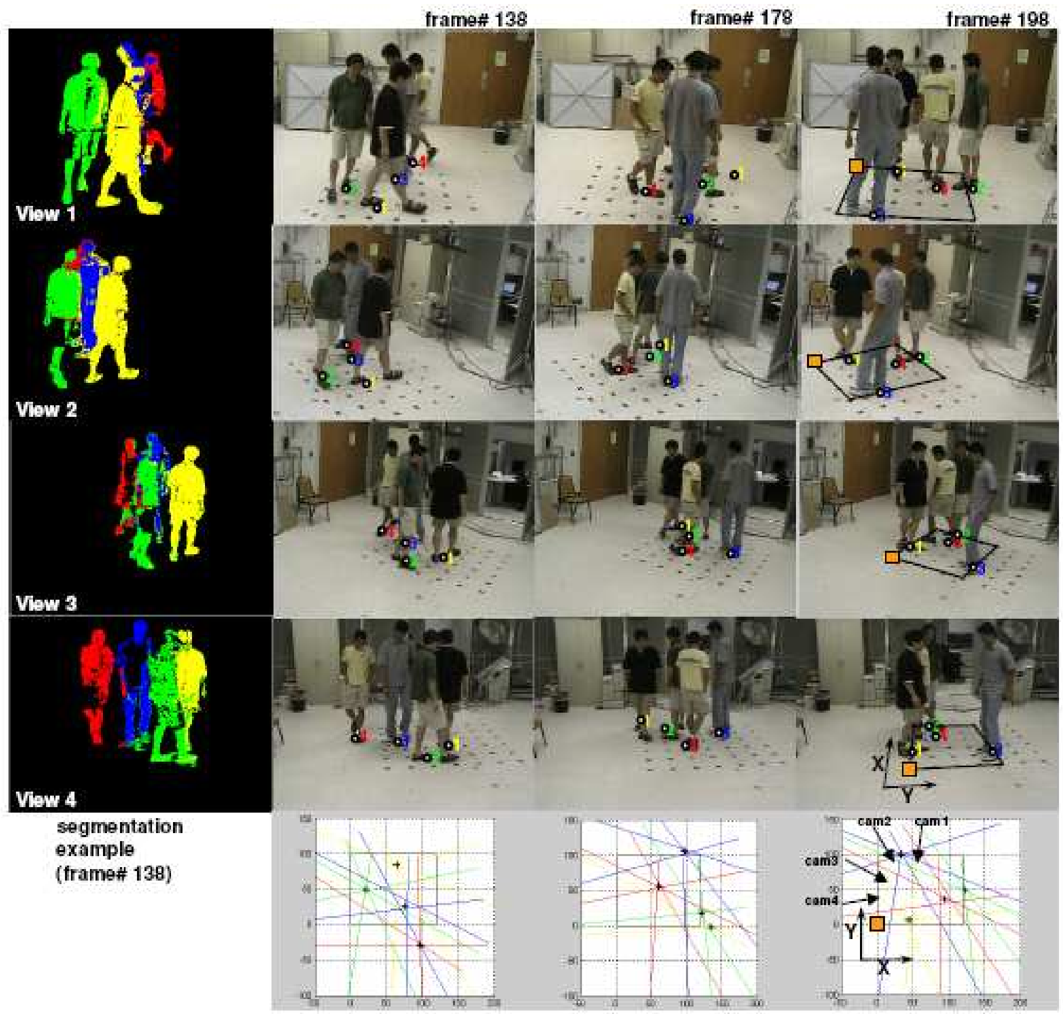


Thresholded and Rectified



Prior Work (Kim and Davis, ECCV 2006)

- Segment at each frame
- Estimate vertical axis.
- Project vertical axes to the reference top down plane.
- Estimate point of intersection.
- *Track location on the ground plane using PFs*.



Overview of Existing Methods

1. Background Subtraction

    Tracking on Individual Image Planes.

2. Data Association across cameras.

    Planar Constraint.

3. Location Estimation on Ground Planes

    Temporal Smoothing.

No explicit modeling of how camera positioning affects results.

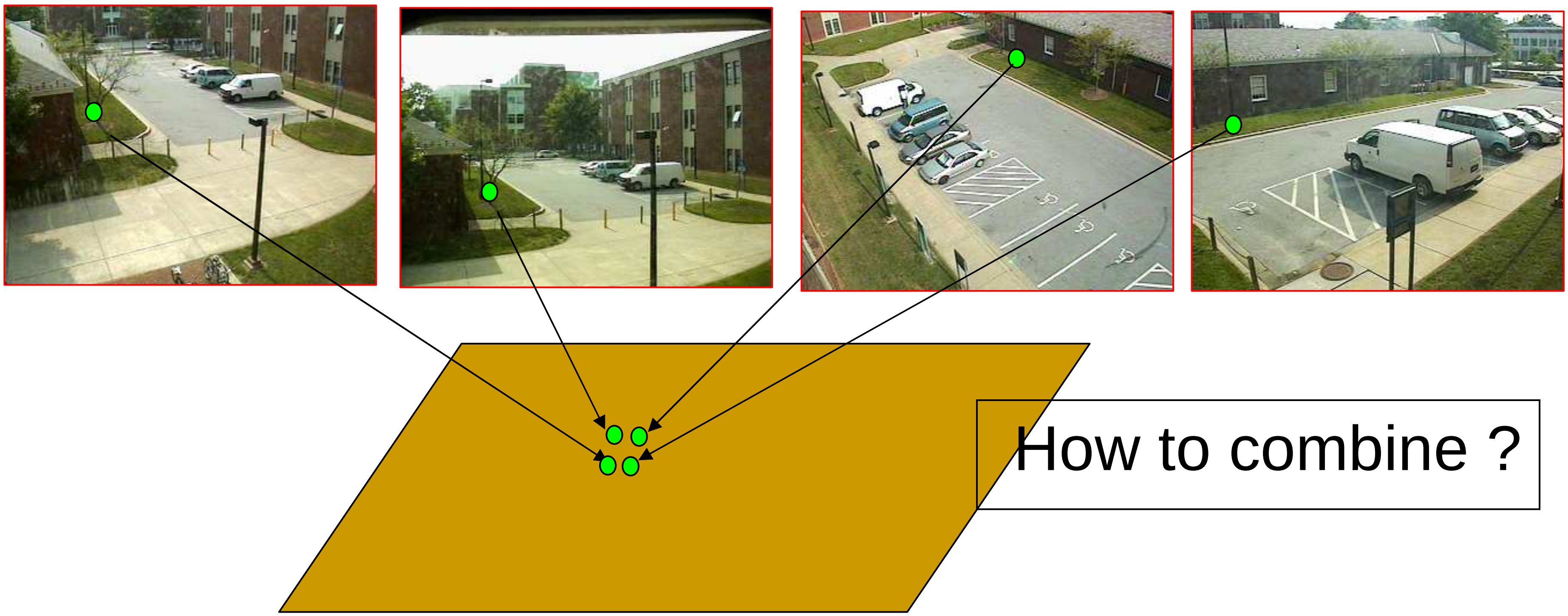
Motivation

- Objects are imaged at different resolutions.

- As a consequence, accuracy of estimation of object location is NOT uniform over the plane and across multiple views.

How does camera positioning affect location estimation ?

Problem definition



- Model the Image Plane location as a random variable (r.v.)
- Study how the distribution of the r.v. changes under the homography.

The transformed RV's statistics would decide the appropriate fusion scheme

Modeling Concerns

- Image Plane to Ground Plane transformation is projective.

Need to study transformation of random variables under projective transformations.

- Non-linearity of the projective transformation.
- Nature of uncertainty on the Image Plane.
  - Imaging (sensor) noise.
  - Estimation process: Kalman, Particle filter ...

Deriving the Distribution

Assume a Gaussian distribution on the image plane

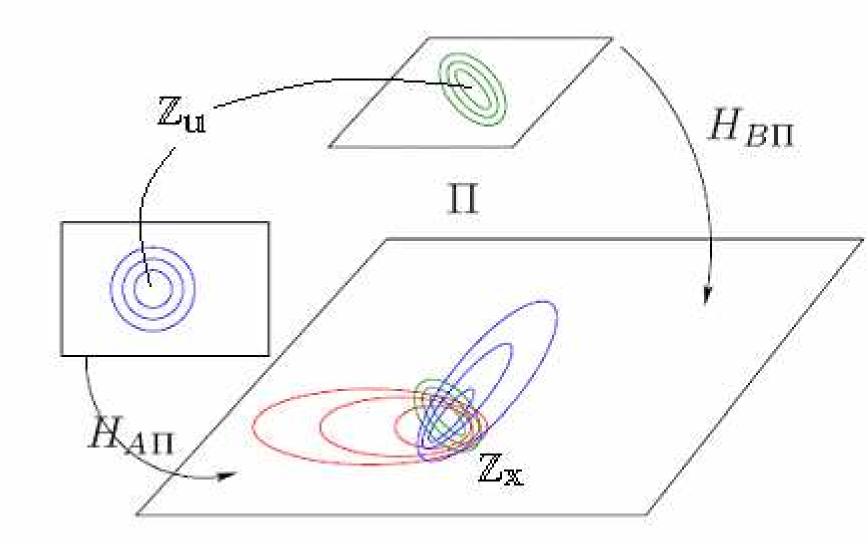
- Z_u is r.v. modeling image plane location

$$H : Z_U \in \mathbb{P}^2 \rightarrow Z_X \in \mathbb{P}^2$$

$$\overline{Z_U} \mapsto \overline{Z_X} = H\,\overline{Z_U}$$

$$Z_X = \begin{bmatrix} Z_x \\ Z_y \end{bmatrix} = \begin{bmatrix} \dfrac{h_1^T \overline{Z_U}}{h_3^T \overline{Z_U}} \\[2ex] \dfrac{h_2^T \overline{Z_U}}{h_3^T \overline{Z_U}} \end{bmatrix}$$

Direct Linear Transformation



- Z_X has a *Ratio of Gaussian* Distribution

Ratio of Gaussians Quick facts (Marsagilia 1965)

- Can be translated to *Ratio of Standard Normals W(a,b).*

X,Y are arbitrary correlated Normals

$$W_1 = \frac{X}{Y} \sim c_0 + c_1 W(a,b),$$

$$W(a,b) = \frac{Z_1 + a}{Z_2 + b}$$

Z1, Z2 are Std. Independent Normals

c0, c1 are constant indep. of *E(X), E(Y)*

$$b = \frac{\mu_Y}{\sigma_Y}$$

Ratio of Gaussians (RoG) Distribution

- The density function of Ratio of Std. Gaussians (Marsagilia 1965, 2006) is of the form:

mixture strength

$$f(t) = e^{-(a^2+b^2)/2} \frac{1}{\pi(1+t^2)} + \left(1 - e^{-(a^2+b^2)/2}\right) f_1(t)$$

Cauchy term

well-behaved distribution

- Implications:
  - Moments do not exist!
  - Sample Median and MLE form possibly useful statistics.

Link to the Line at infinity.

- Projective Transformation can be factored into similarity, affine and projective components.

$$H = H_A H_P = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix}$$

$$Z_x = \frac{h_1^T \overline{Z_U}}{h_3^T \overline{Z_U}}, Z_y = \frac{h_2^T \overline{Z_U}}{h_3^T \overline{Z_U}}$$

- Implies $\boldsymbol{h_3}$ is proportional the projection of the line at infinity.
- Further, in W(*a*,*b*), $b = \frac{\mu_Y}{\sigma_Y} = \frac{h_3^T \mu_U}{\sigma_Y} \propto \frac{L_\infty^T \mu_U}{\sigma_Y}$
- Finally, mixture strength is $e^{-(a^2+b^2)/2}$

Lemma: Strength of Cauchy Component

The strength of the Cauchy component in the distribution depends on the distance of the true location of imaged point from the projection of the line at infinity.

If the imaged region is sufficiently far-far away from the projection of the Line at Infinity, then the strength of the Cauchy component is negligible.

Further, when the Cauchy component is of negligible strength, "pseudo"-moments can be computed and the overall distribution is approximated to a high accuracy with a Normal Density [Marsagilia, 1965].

Projective Transformations under Affine Approx.

- Normal → Normal can be obtained with an affine transformation.
- However, this mapping is only POINTWISE, and does not extend to regions.
- However, a local approximation is still valid (provided the imaging is sufficiently far away from the Line at Infinity).

Degenerate Cases

- Affine camera or principal axis parallel to plane normal.
  - Homography is an affine transformation.
  - *The strength of Cauchy component is zero.*
  - No non-linearity in the transformation.

Computing Moments using Approximations

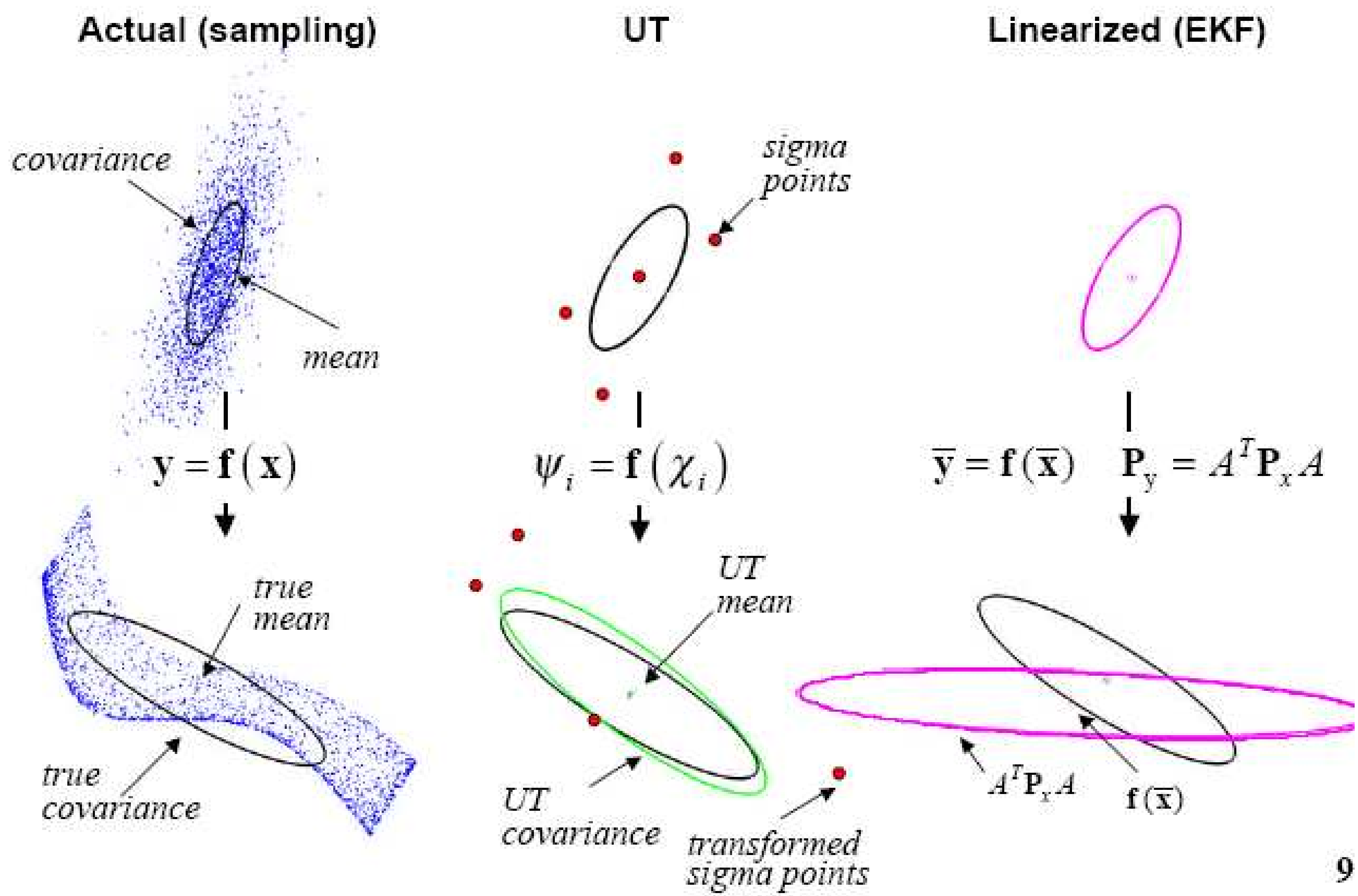
- Linearization
  - First order approximation.

$$Z_X \approx x_0 + \frac{1}{h_3{}^T m_0} J_H(m_0)(Z_U - m_0)$$

  - Variance increases as target approaches horizon.
- Unscented Transformation (Julier 1996)
  - Propagate moments across any non-linear transformation.
  - A second order approximation.

Combine first two moments from each camera using the min. variance estimator.

Unscented Transformation



(http://cslu.cse.ogi.edu/publicatio

Algorithm

- Compute Homography between each view and plane.
- Start with Image plane moments at each camera.
- Use UT to obtain moments of the RV over the plane.
- Min. Variance Estimator to fuse.
- *Dynamical System*: The observation model takes the role of the Min. Var estimator.
  - State Model: Constant Velocity

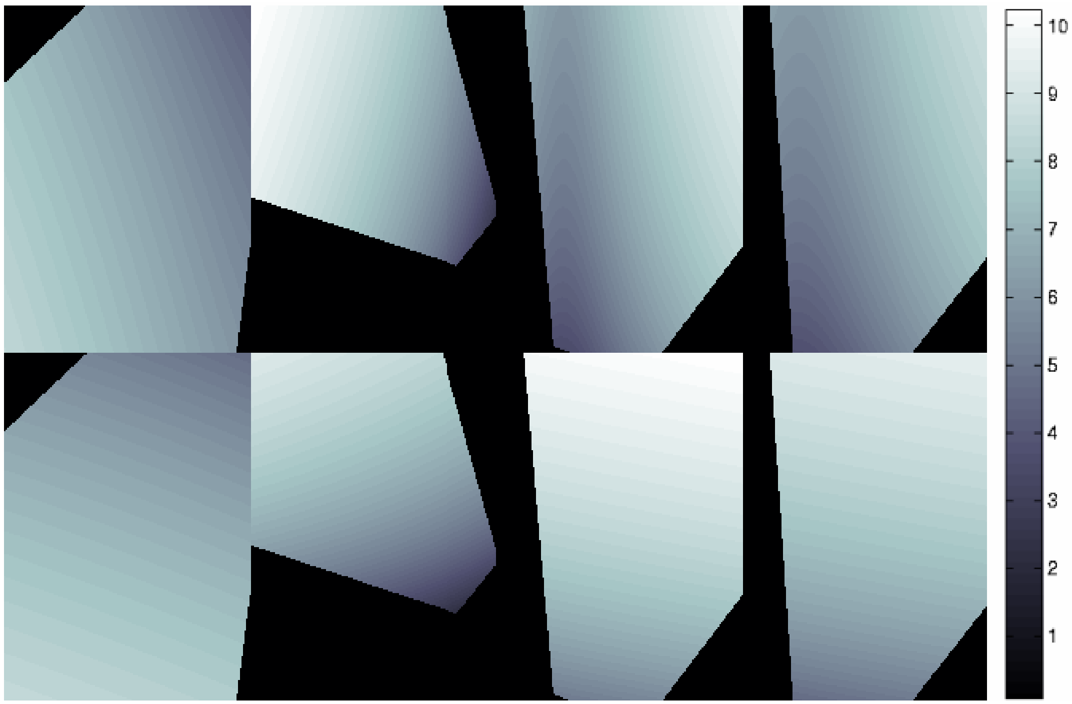
$$\mathbf{x}_t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \omega_t$$

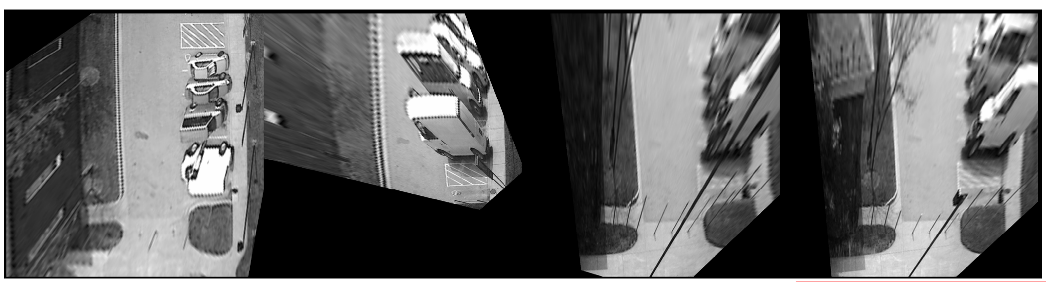
  - Observation Model: Use variance models from the UT.

$$\mathbf{y}_t = \begin{bmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_M \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & & & \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_t + \Omega_t$$

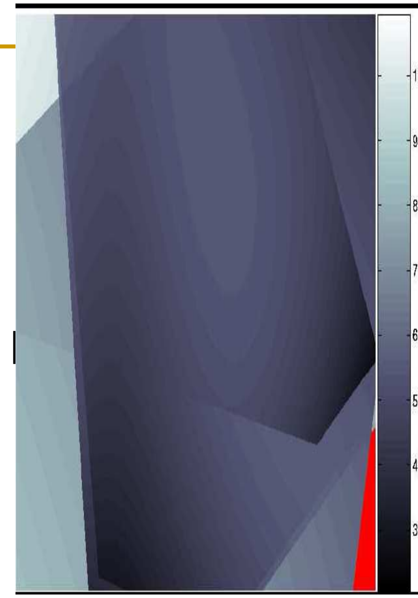
Tracking result

Experiment (Location Estimation)





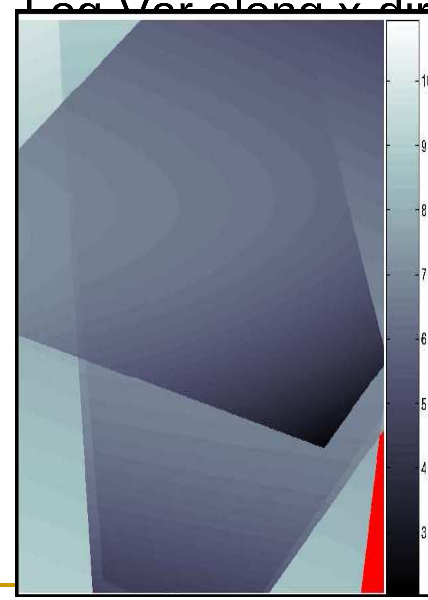
Camera Vi



ection

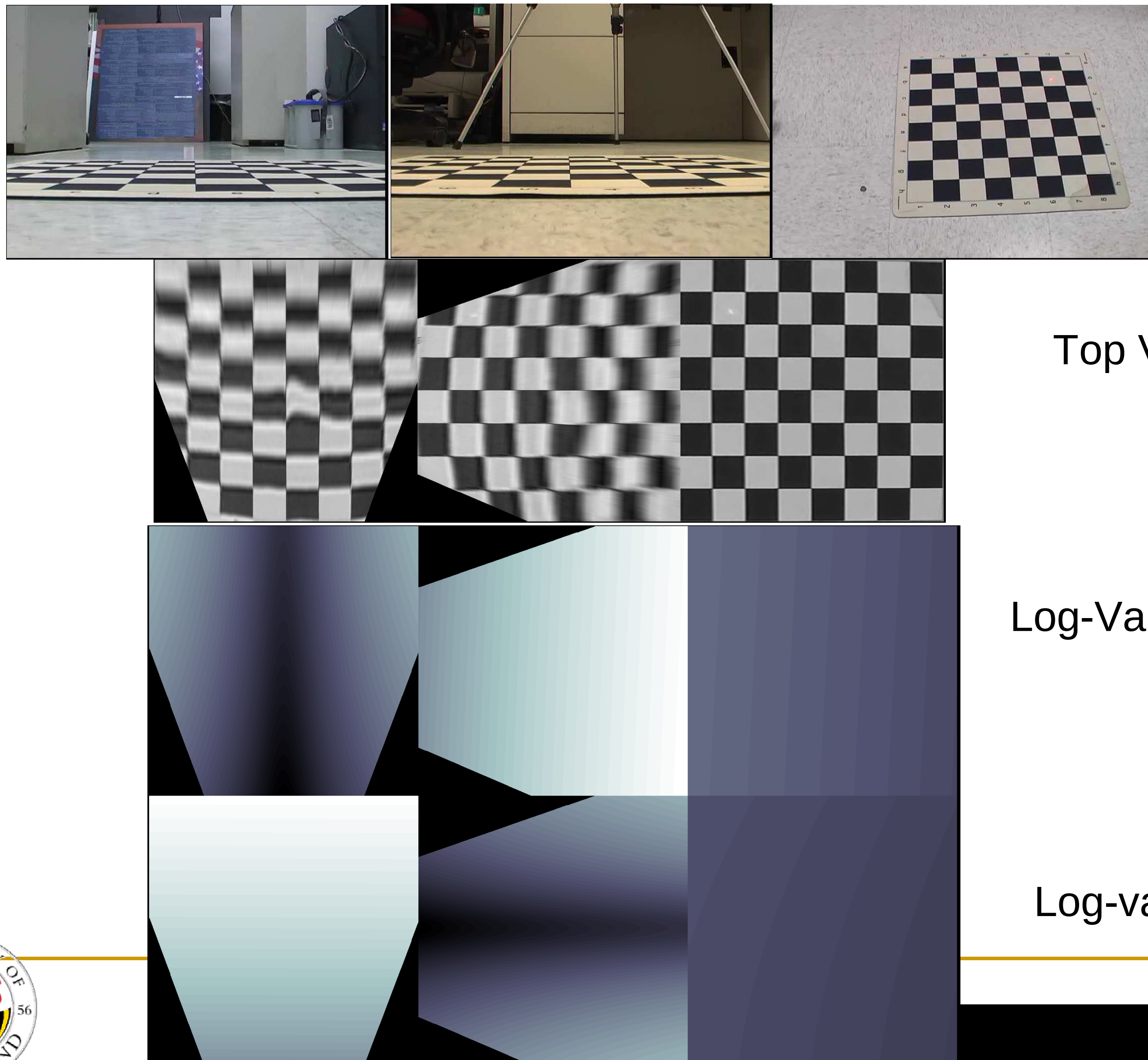
Log Var. of Min. Variance estimator along x-direction

Log Var along x-direction



Log Var. of Min. Variance

Multi-Camera Tracking on a Plane



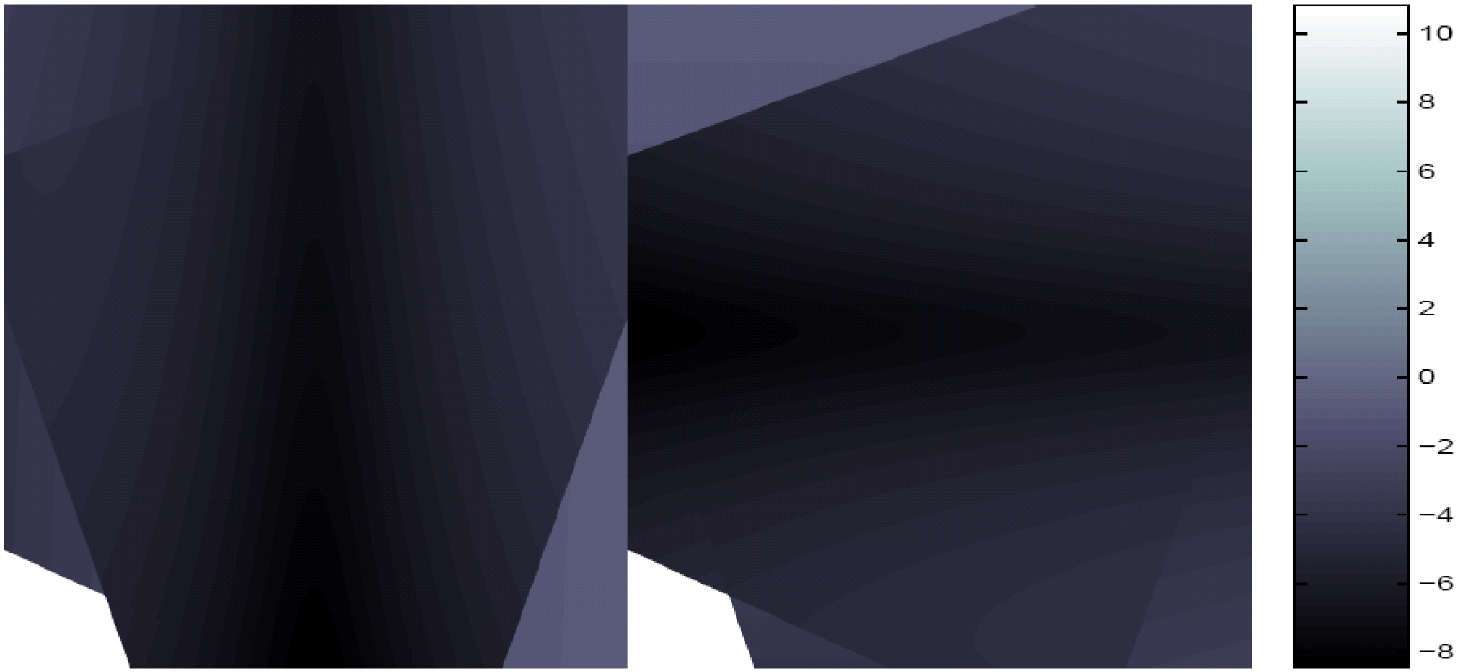
Camera View

Top View

Log-Var along x-direction

Log-var along y-direction

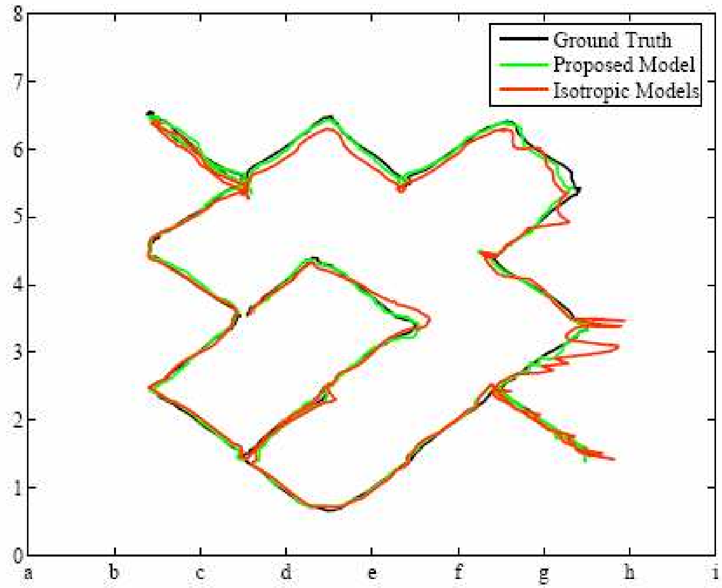
Multi-Camera Tracking



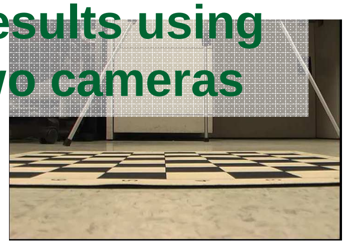
Log Var. of Min. Variance estimator along x-direction

Log Var. of Min. Variance estimator along y-direction

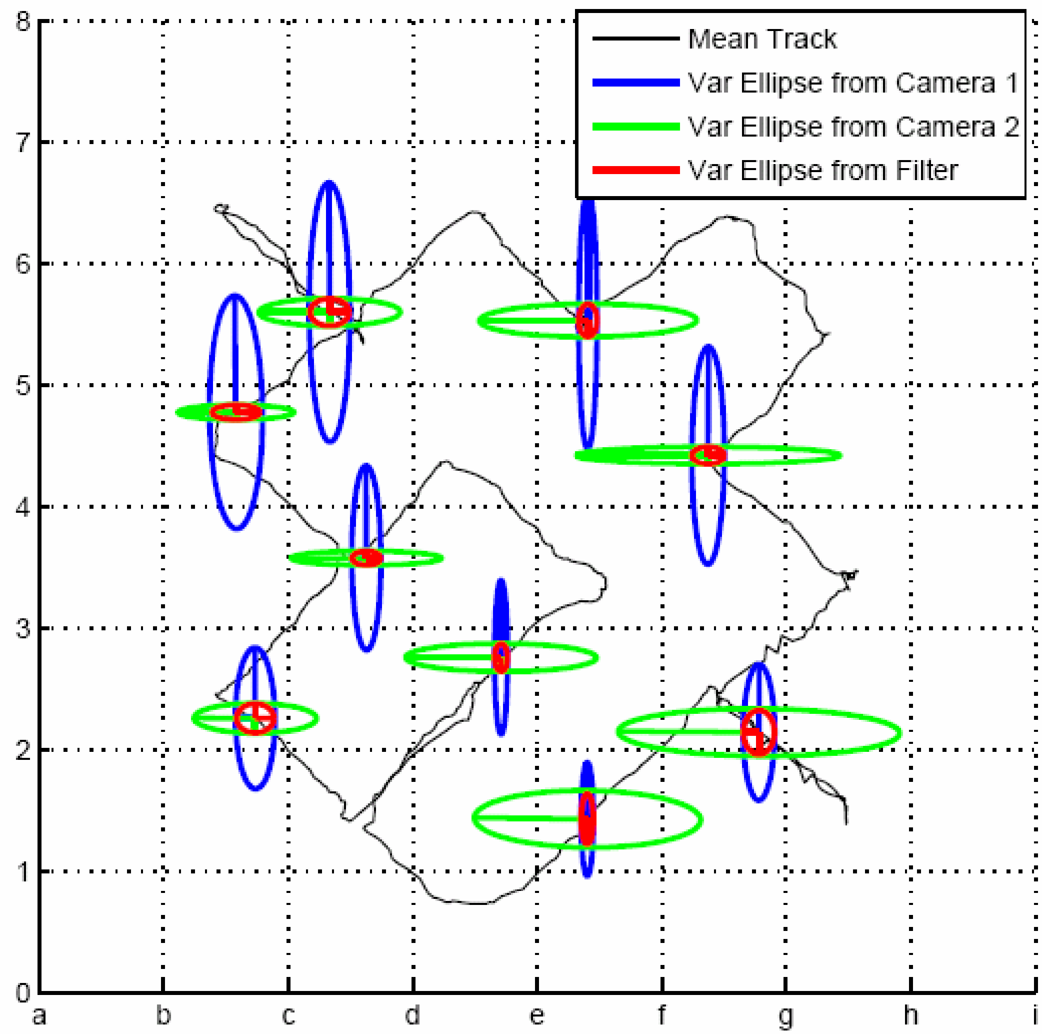
Multi-Camera Tracking



Tracking Results using only the two cameras



Used for Ground Truth

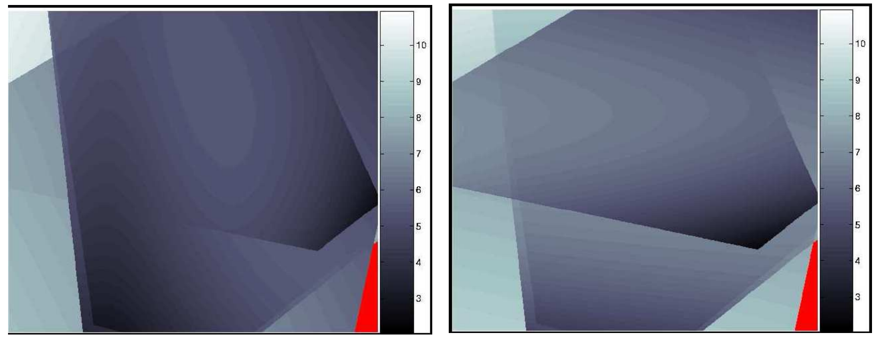


Summary

- Need to model camera-plane dependence for multi-view fusion.
- Projective transforms Normal to Normal ONLY when the region of interest is imaged sufficiently away from the Line at Infinity at each view.
- Using Unscented Transformations to estimate moments, a minimum variance estimator is designed to fuse multi-view estimates.

Other Potential Use of such Modeling

- Camera Placement



- Stabilization/Mosaic
  - Fusing Gradients.