Structure From Motion without Correspondence

Frank Dellaert

Carnegie Mellon University

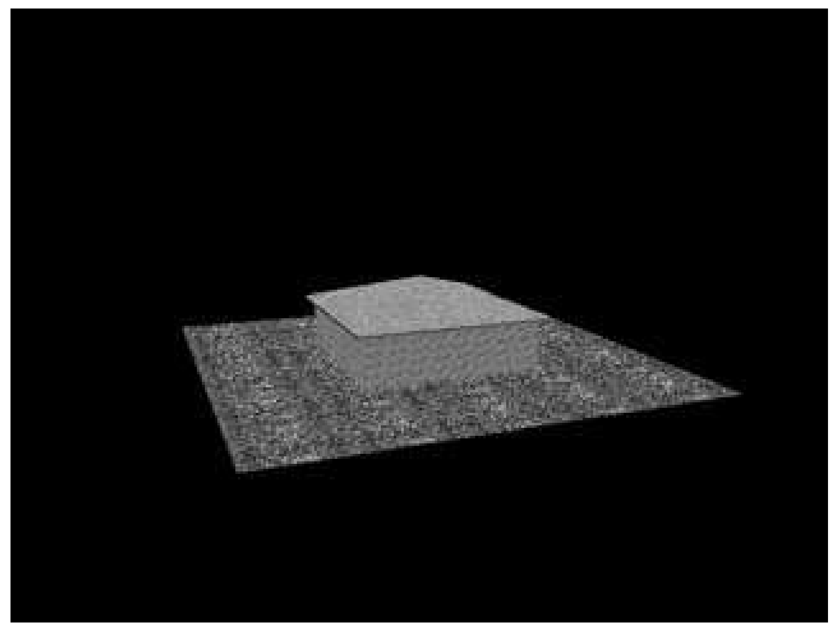
Unrestricted Scene Modeling from Sets of Images

Frank Dellaert

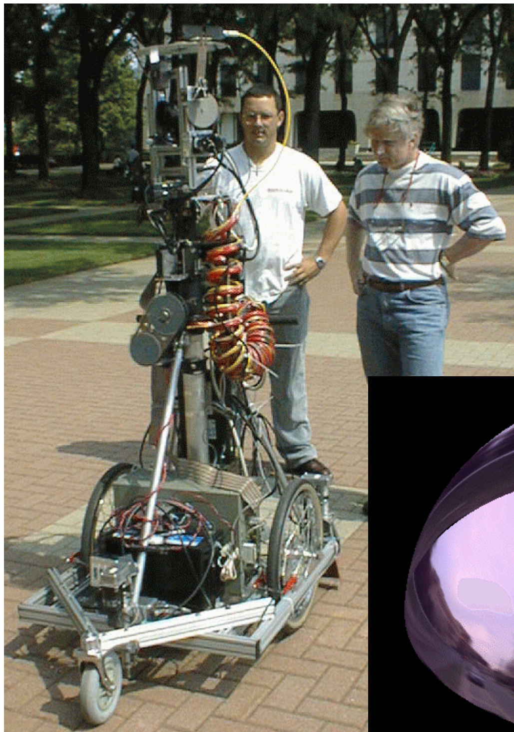
Carnegie Mellon University

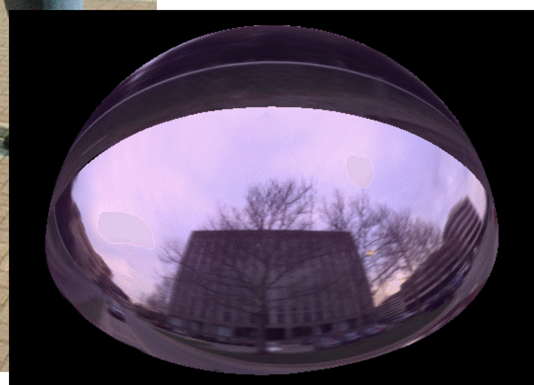
3D Structure from Motion

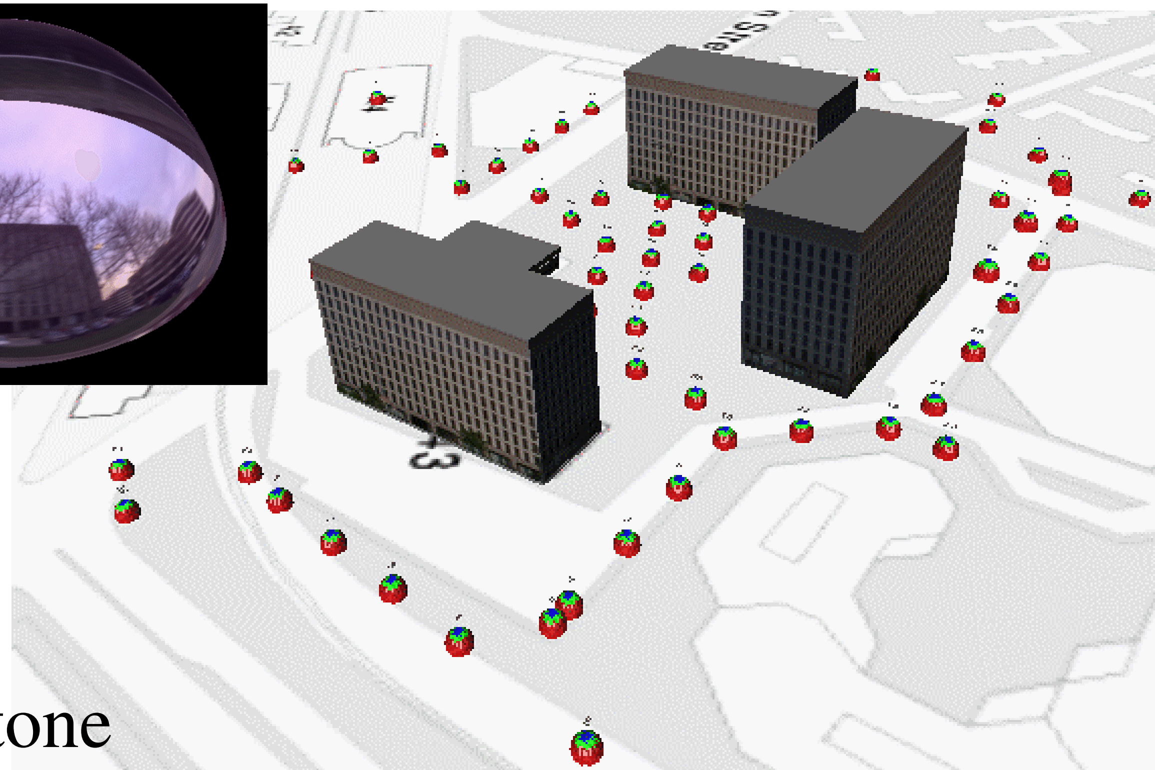
- Estimate 3D Structure from 2D Images
- E.g. CMU Virtualized Reality (Kanade)



MIT City Scanning Project







Seth Teller, Matt Antone

Why ?

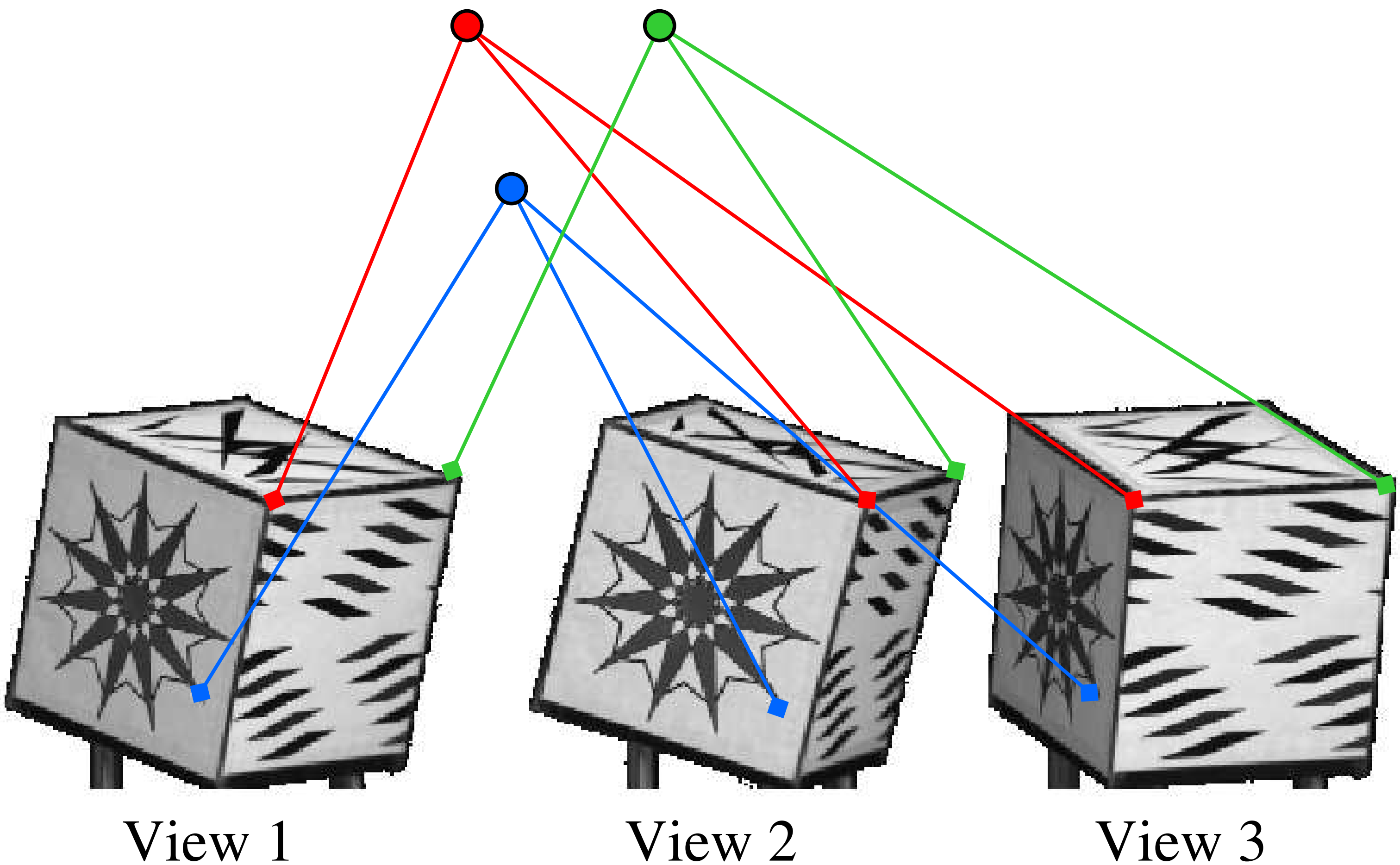
- The World is three-dimensional
  - Acting in the world -> need 3D
  - Simulating the world -> need 3D
  - Investigating the world -> need 3D
- Tons of applications

Really 2 Problems !

Correspondence

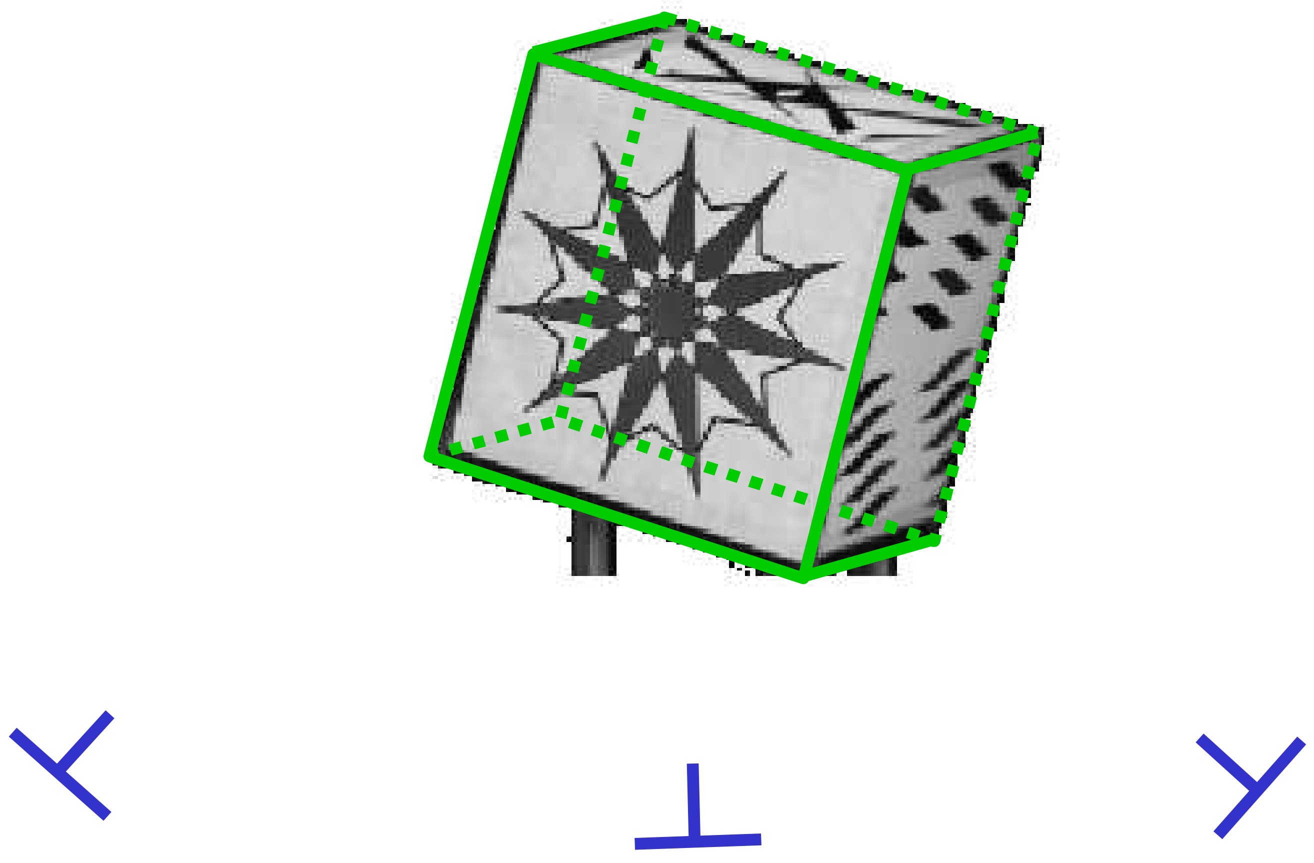
Optimization

Correspondence



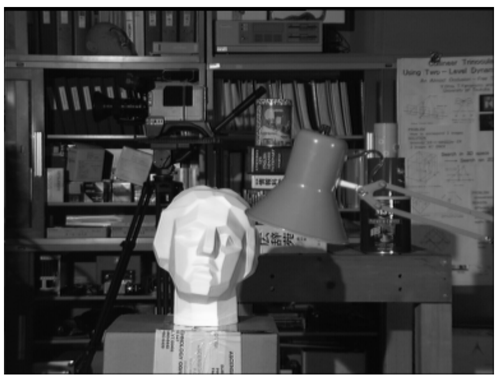
Optimization

- Most Likely 3D Structure and "Motion" given Correspondence:

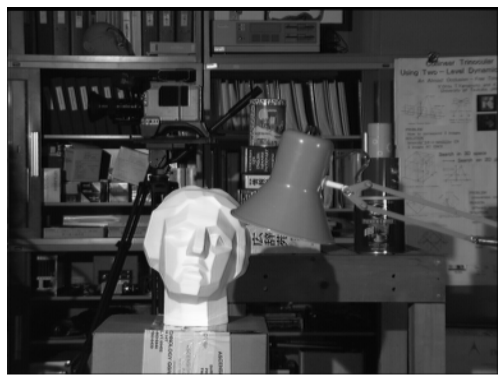


Dense Correspondence

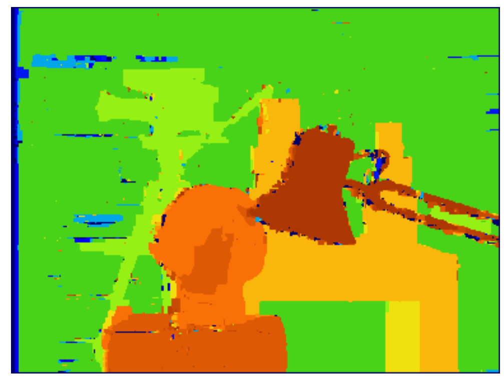
- Stereo -> depth map (Marr)
- Depth from Sequence (Aloimonos, Sarnoff…)



Left



Right

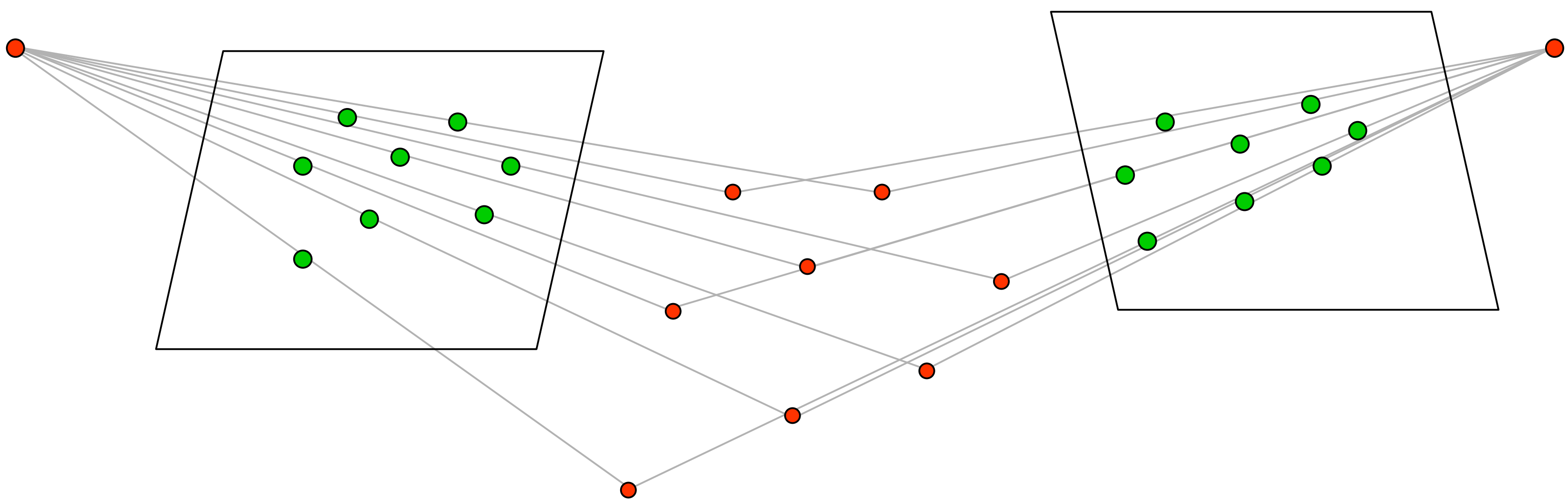


Depth Map

Boykov, Veksler & Zabih 98

Sparse Correspondence

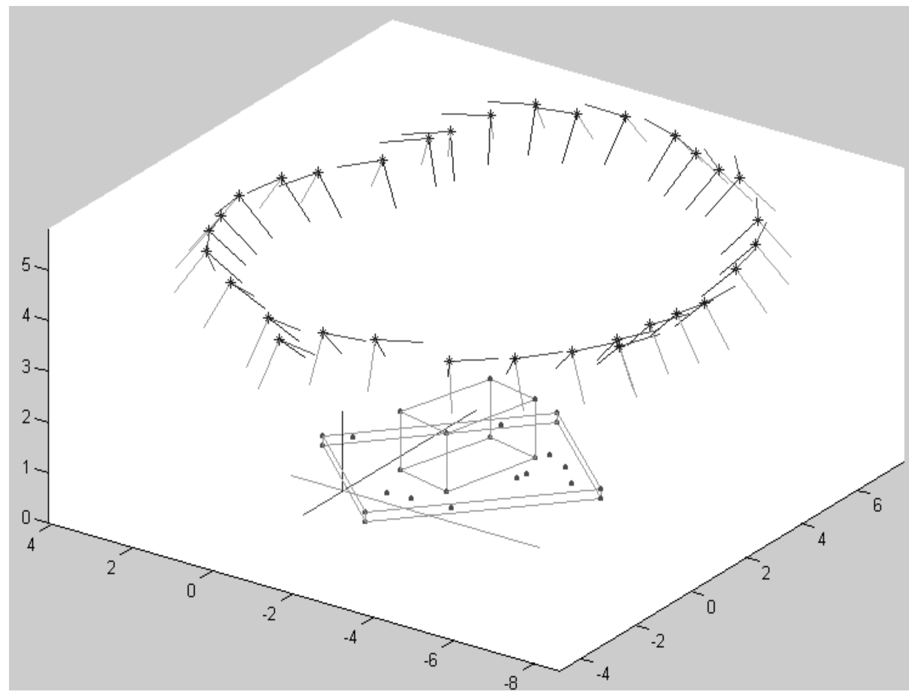
- Longuet-Higgins 1981: 8 point algorithm



Optimization

- Most Likely Structure given Correspondence: is Non-linear Least-Squares

$$-\log L(\Theta; \mathbf{U}, \mathbf{J}) = \frac{1}{2\sigma^2} \sum_{i=1}^{m} \sum_{k=1}^{K_i} \|\mathbf{u}_{ik} - \mathbf{h}(\mathbf{m}_i, \mathbf{x}_{\mathbf{j}_{ik}})\|^2$$

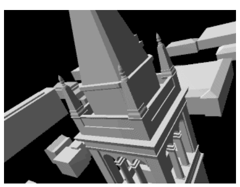


SFM Optimization Literature

- Bundle Adjustment (entire field of photogrammetry, Hartley)

- Orthographic Projection: factorization approach (Tomasi & Kanade)

- Weak perspective factorization (Poelman)



- Model-based constraints (Taylor, Debevec, & Malik)

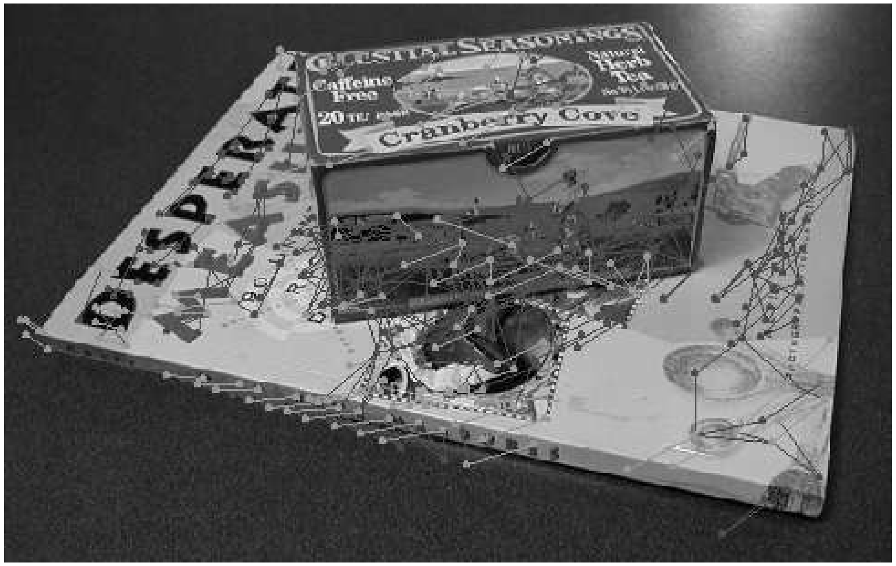
- Sequences: Linear methods based on multi-view constraints (Beardsley, Pollefeys…)

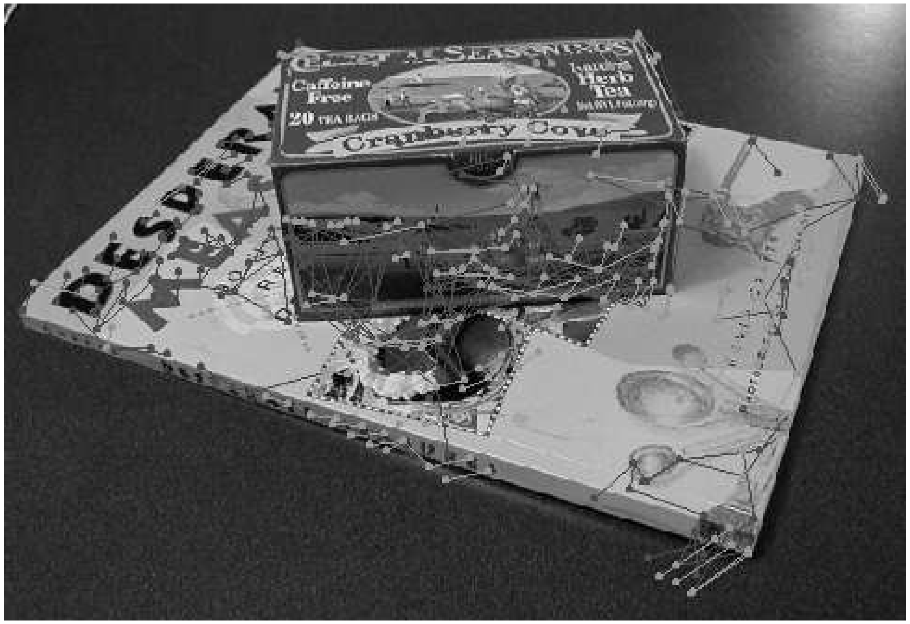
Tracking Features

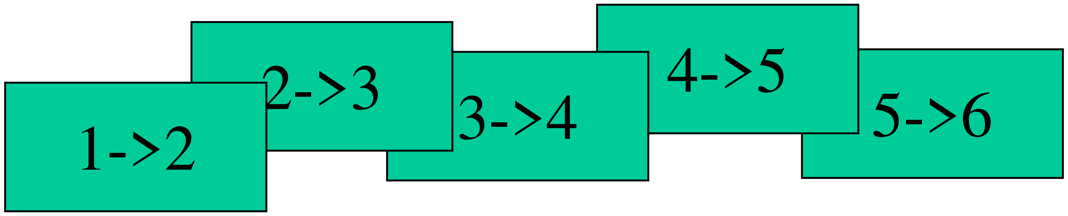
Lucas-Kanade Tracker (figures from Tommasini et. al. 98)

Guessing + Outlier Rejection

- RANSAC (Random Sample Consensus)







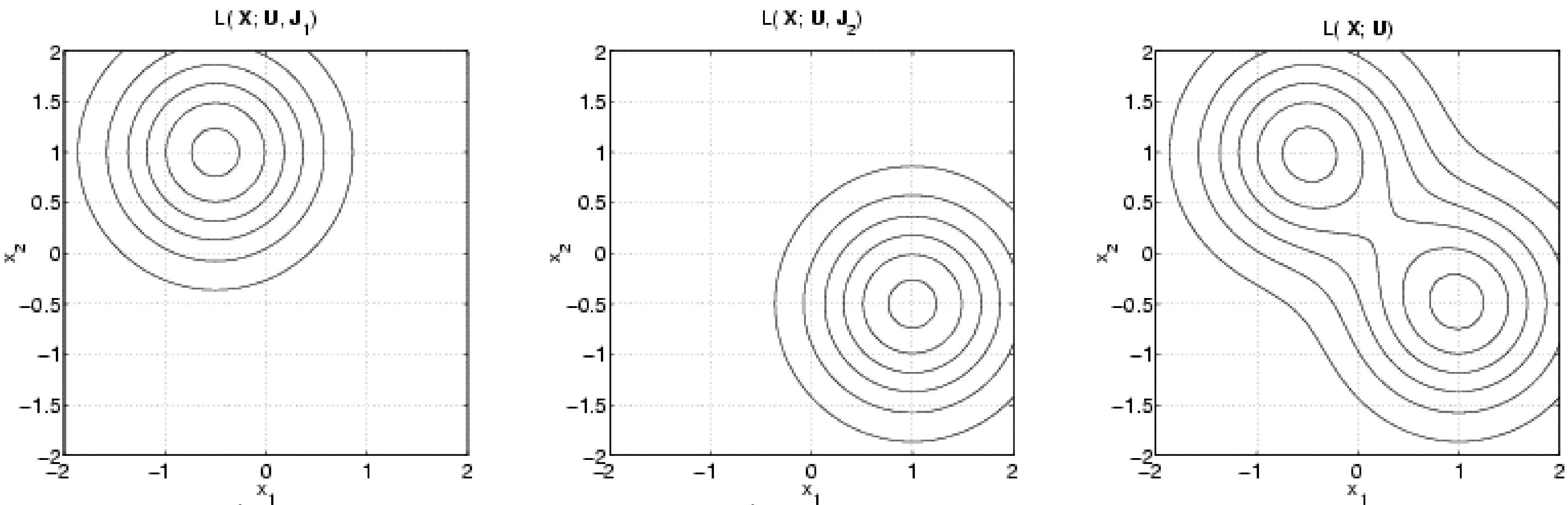
Hartley & Zisserman: “still to some extent a black art”

Today's Talk

ML Structure from Motion
Given NO Correspondence

Assuming (a) sparse (b) no occlusion (today)

Likelihood with Unknown Correspondence



Universe J_1 Universe J_2

$$L(\Theta; \mathbf{U}) = \sum_{\mathbf{J}} L(\Theta; \mathbf{U}, \mathbf{J})$$

Combinatorial Explosion

- In general, #J is combinatorial

- 3 images, 5 features: $5!^3$=1,728,000
- 5 images, 30 features: $30!^5$=1.3131e+162

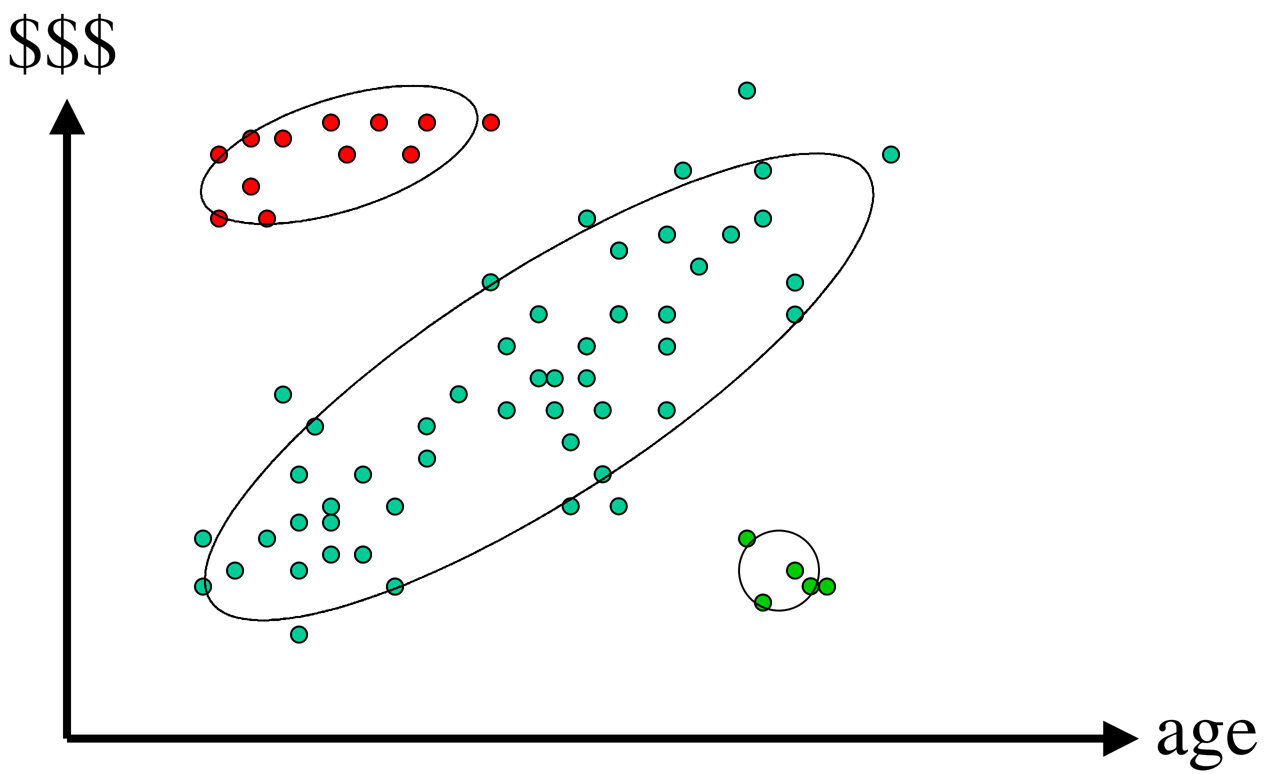
- Total Likelihood = intractable !

Idea 1

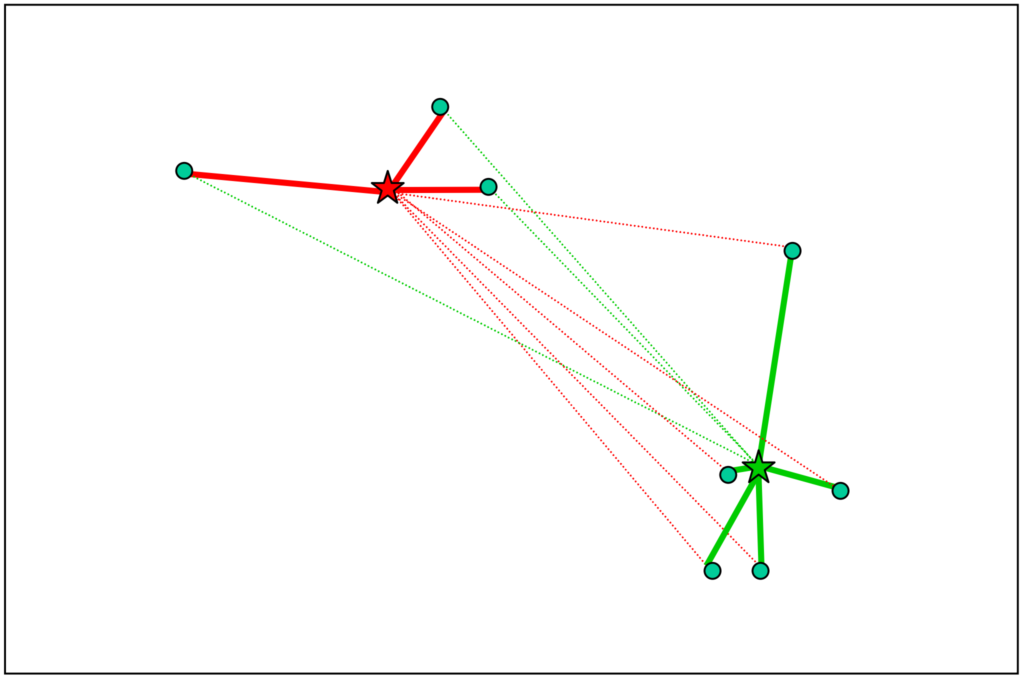
- Analogy with Clustering…

Correspondence as Hidden Data

Clustering



Clustering using the Expectation Maximization Algorithm (EM)



Expectation Maximization

1. **E-step:** Calculate the expected log likelihood $Q^t(\Theta)$:

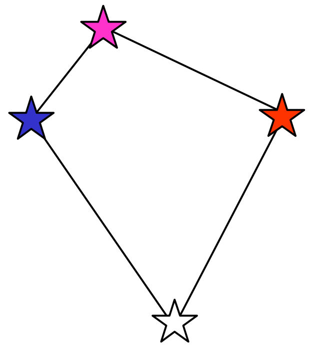
$$Q^t(\Theta) = \sum_{\mathbf{J}} f^t(\mathbf{J}) \log L(\Theta; \mathbf{U}, \mathbf{J})$$

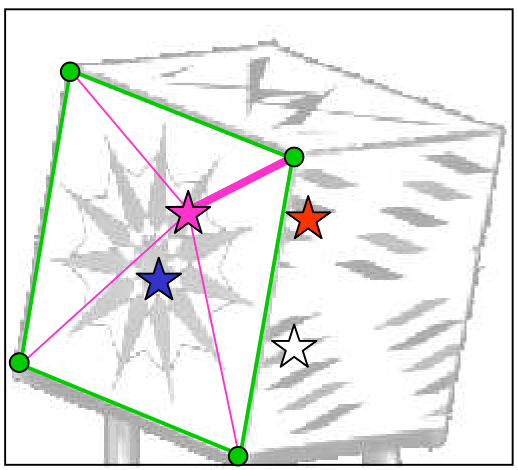
2. **M-step:** Find the ML estimate Θ^{t+1} for structure and motion, by maximizing $Q^t(\Theta)$:

$$\Theta^{t+1} = \underset{\Theta}{\operatorname{argmax}} \; Q^t(\Theta)$$

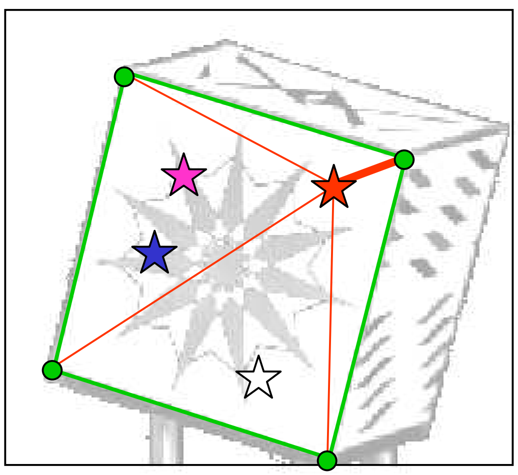
•Expectation taken wrpt hidden assignments: $f^t \triangleq P(\mathbf{J}|\mathbf{U}, \Theta^t)$.

Correspondence Weights

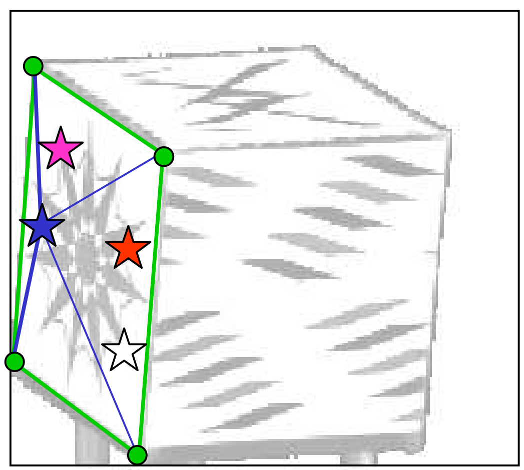




View 1



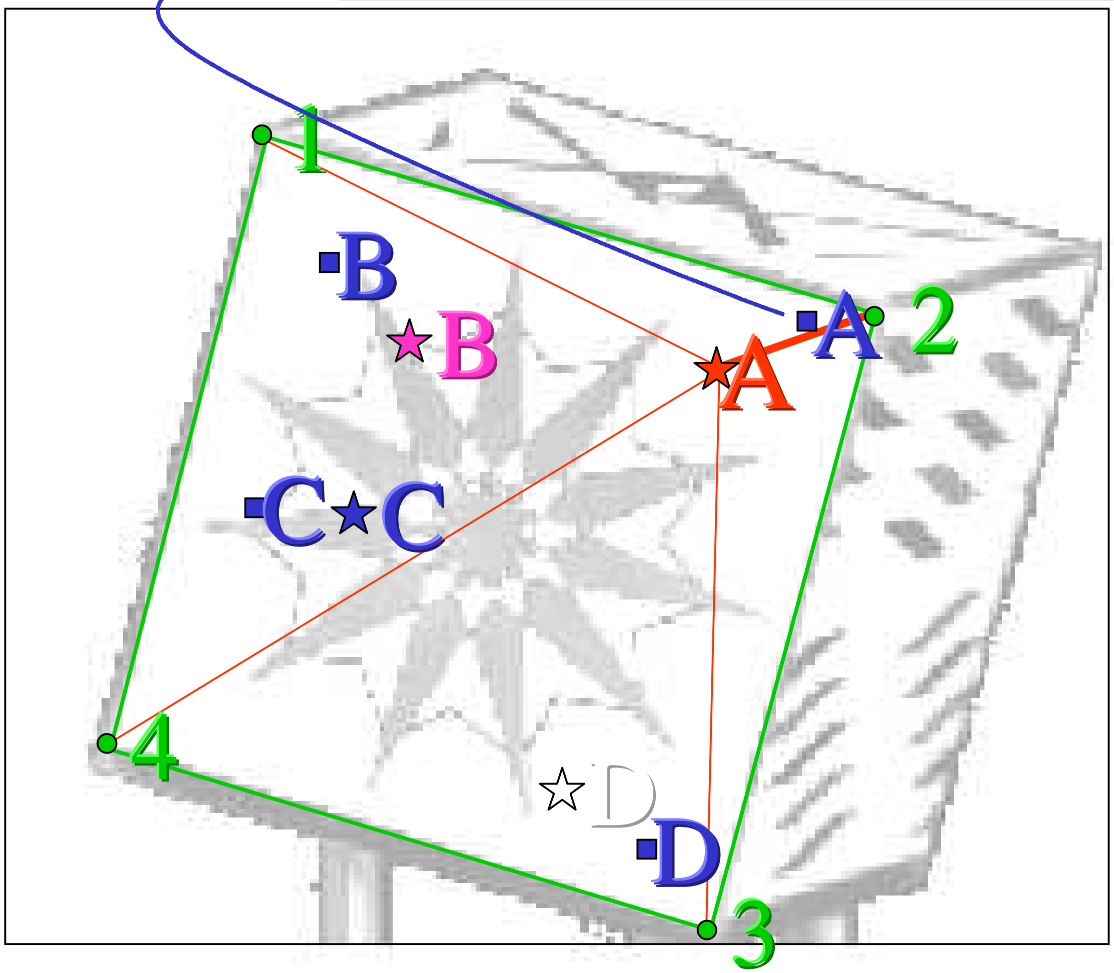
View 2



View 3

Virtual Measurements

$$v_A = f_{1A}u_1 + f_{2A}u_2 + f_{3A}u_3 + f_{4A}u_4$$



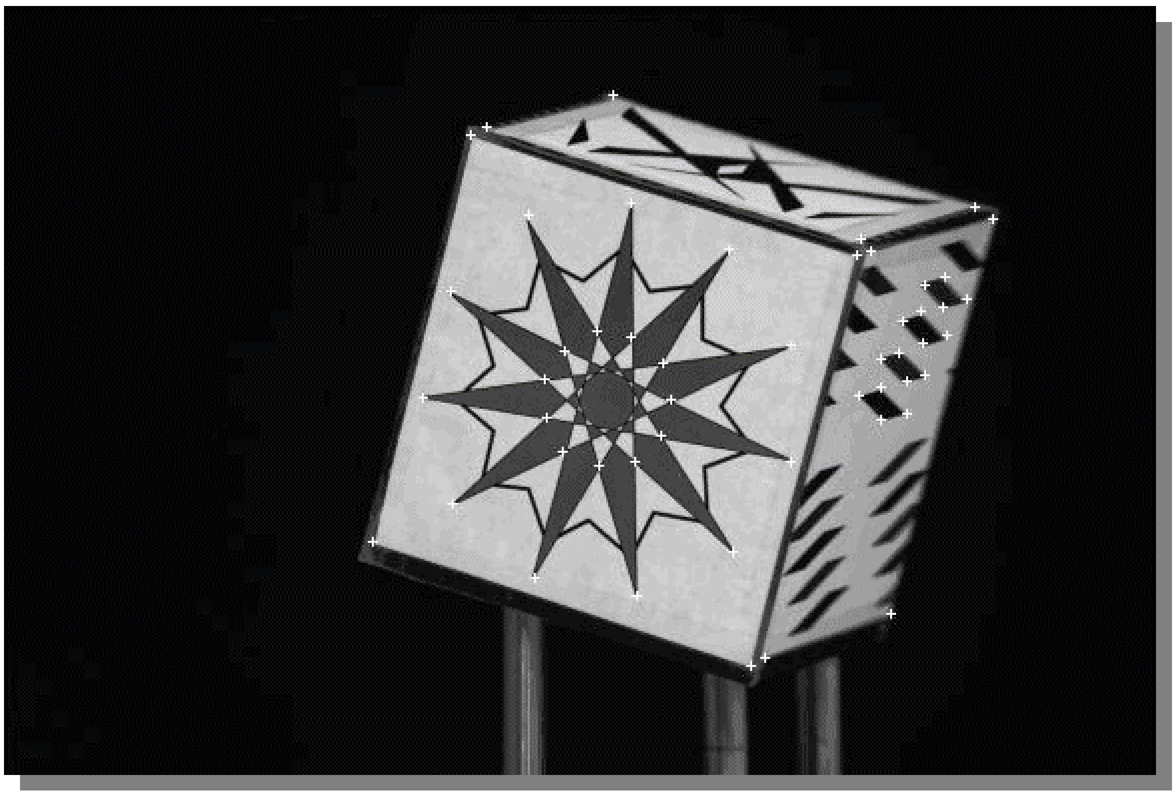
View 2

Summary Algorithm:

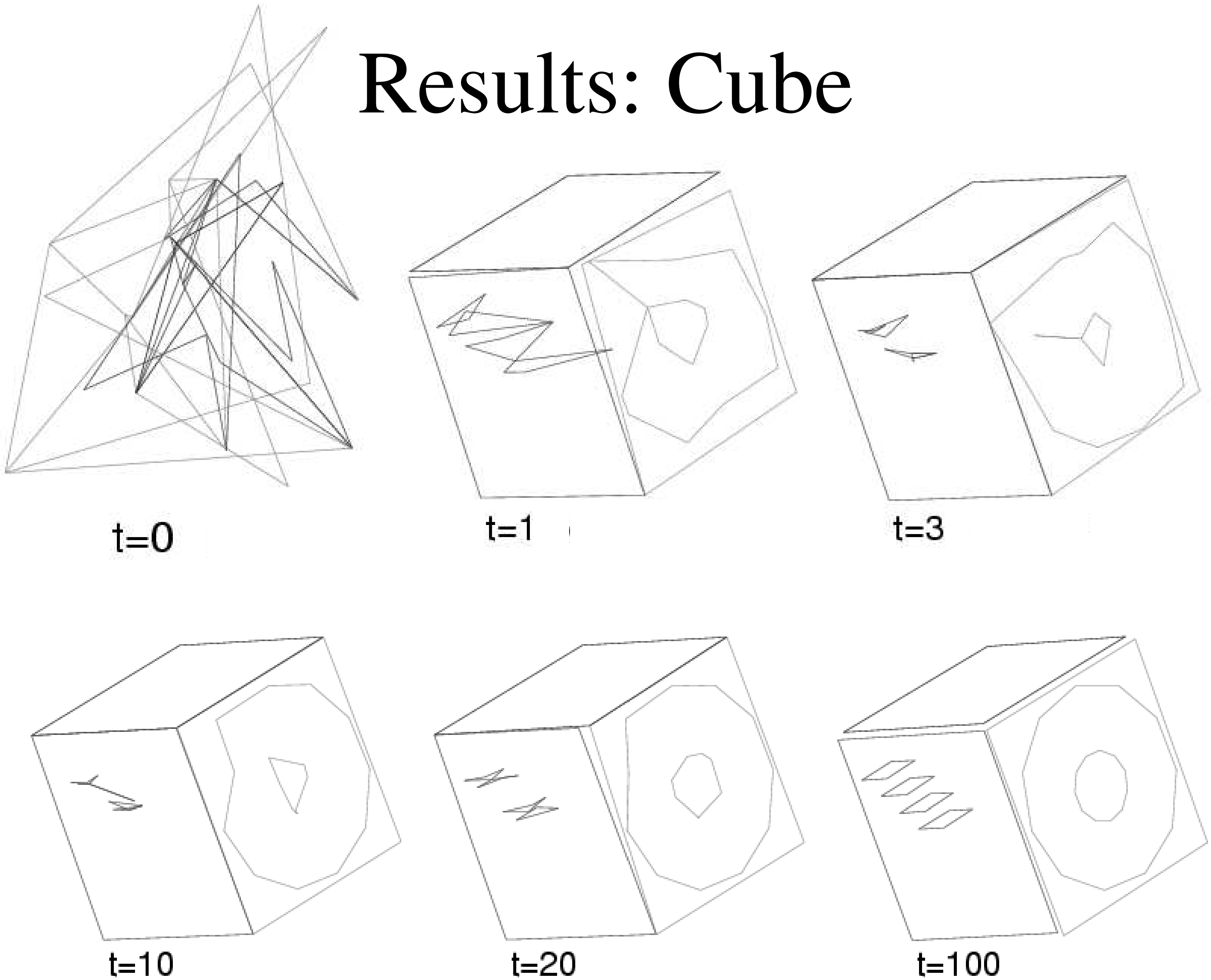
1. Generate an initial structure and motion estimate Θ^0.
2. In each image, calculate the n^2 weights f_{ijk}
3. Calculate the virtual measurements $\mathbf{v}_{ij}^t$.
4. Find the new estimate Θ^{t+1} for structure and motion using the virtual measurements $\mathbf{v}_{ij}^t$.
5. If not converged, return to step 2.

Cube

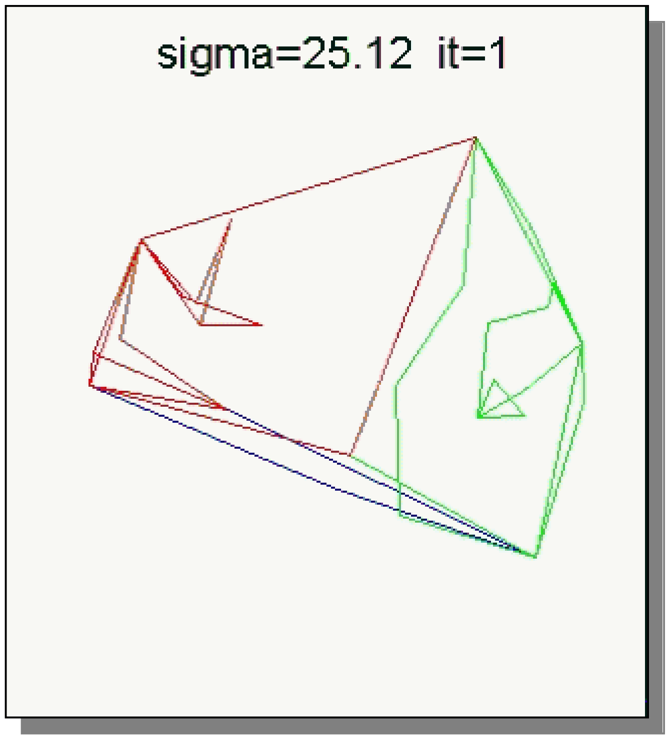
(images presented in random order)



Results: Cube



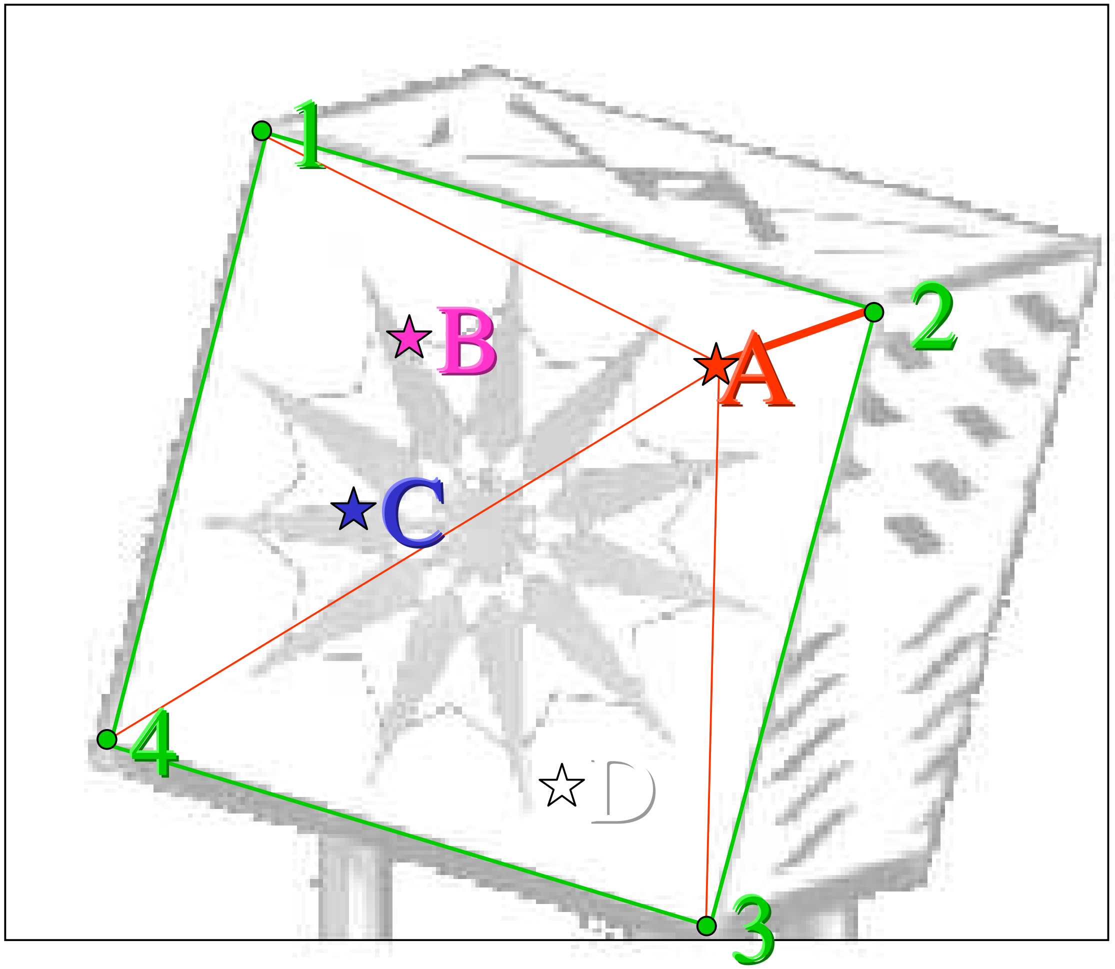
Results Cube



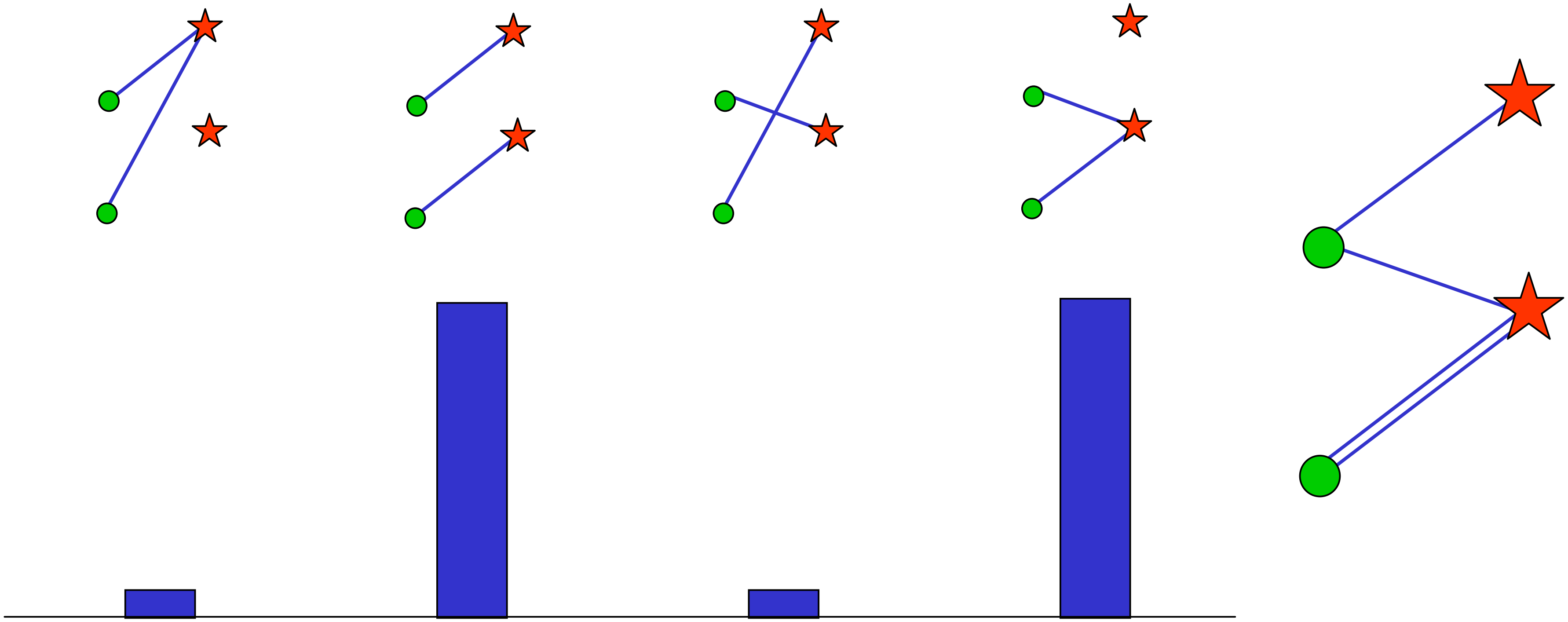
Part II

The Evil Mutual Exclusion Constraint

Calculating weights = key !

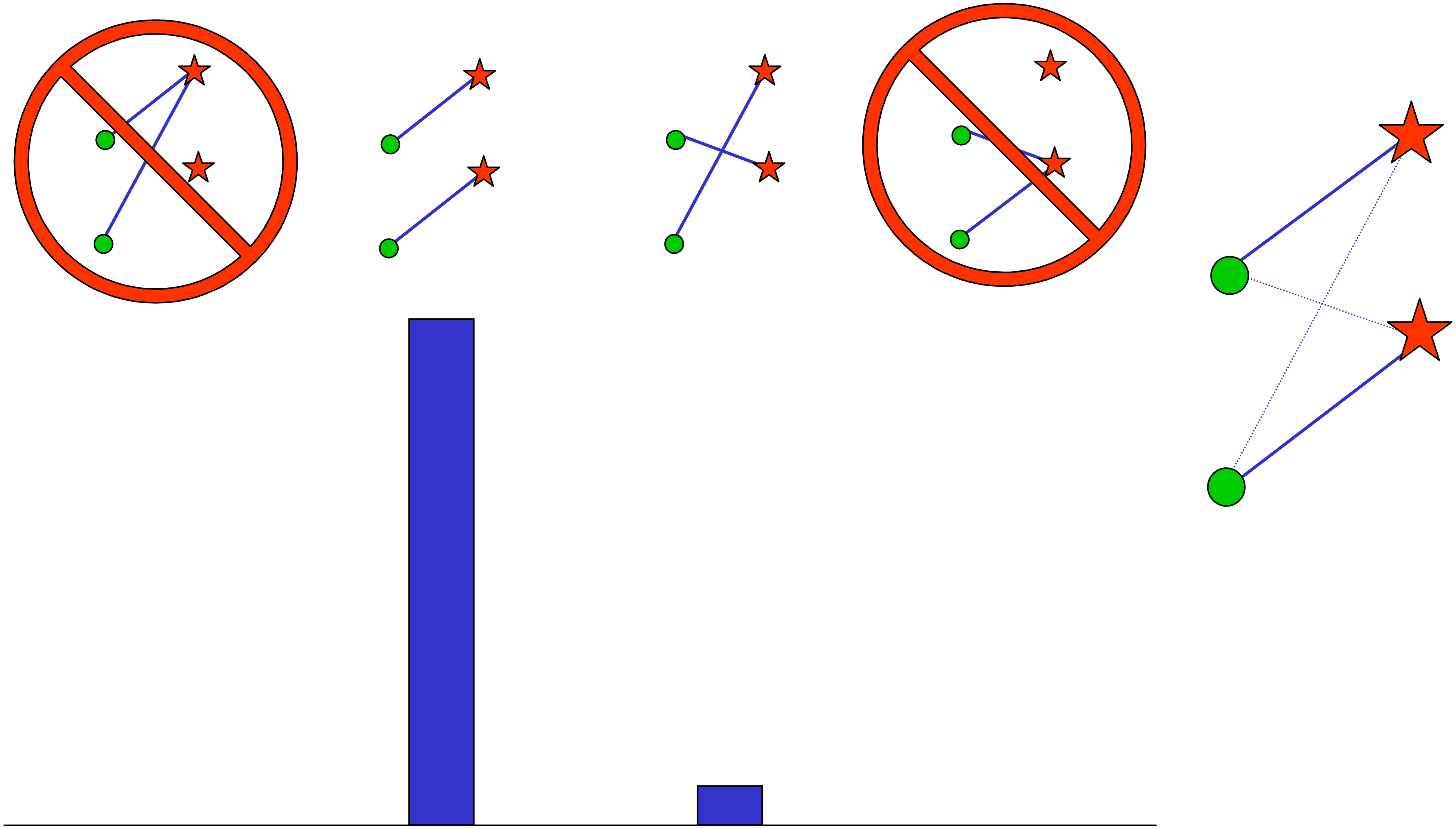


Calculating Weights



Weights = Marginal Probabilities !

Mutual Exclusion



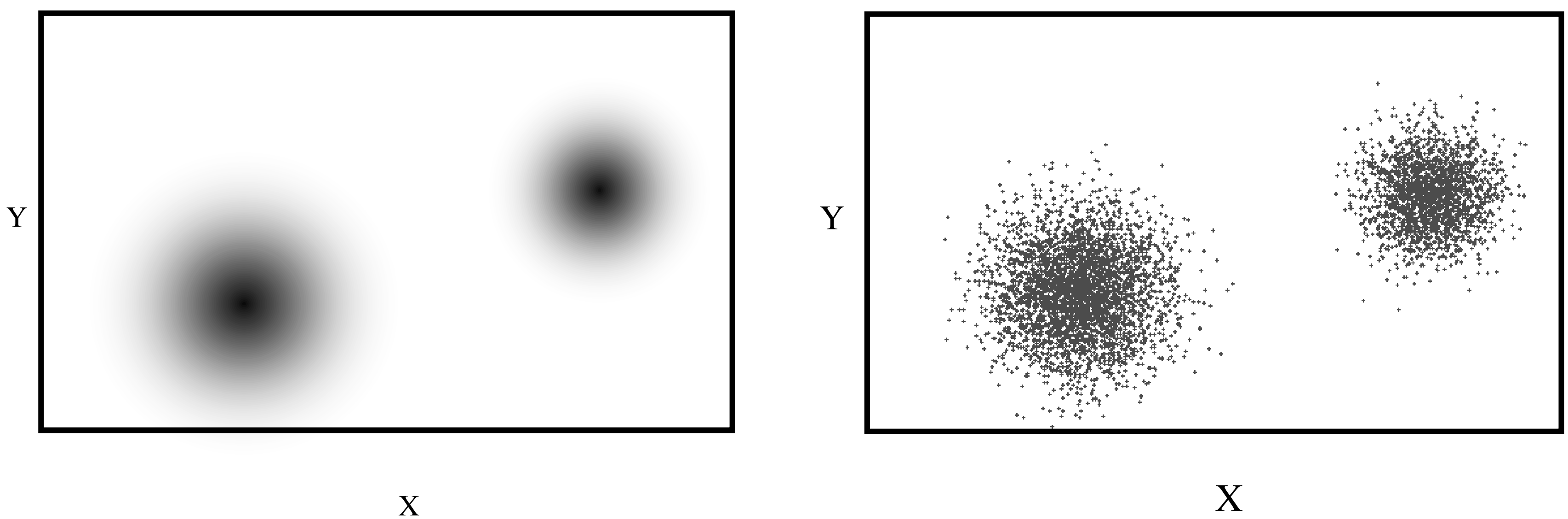
Idea 2

- Sample !

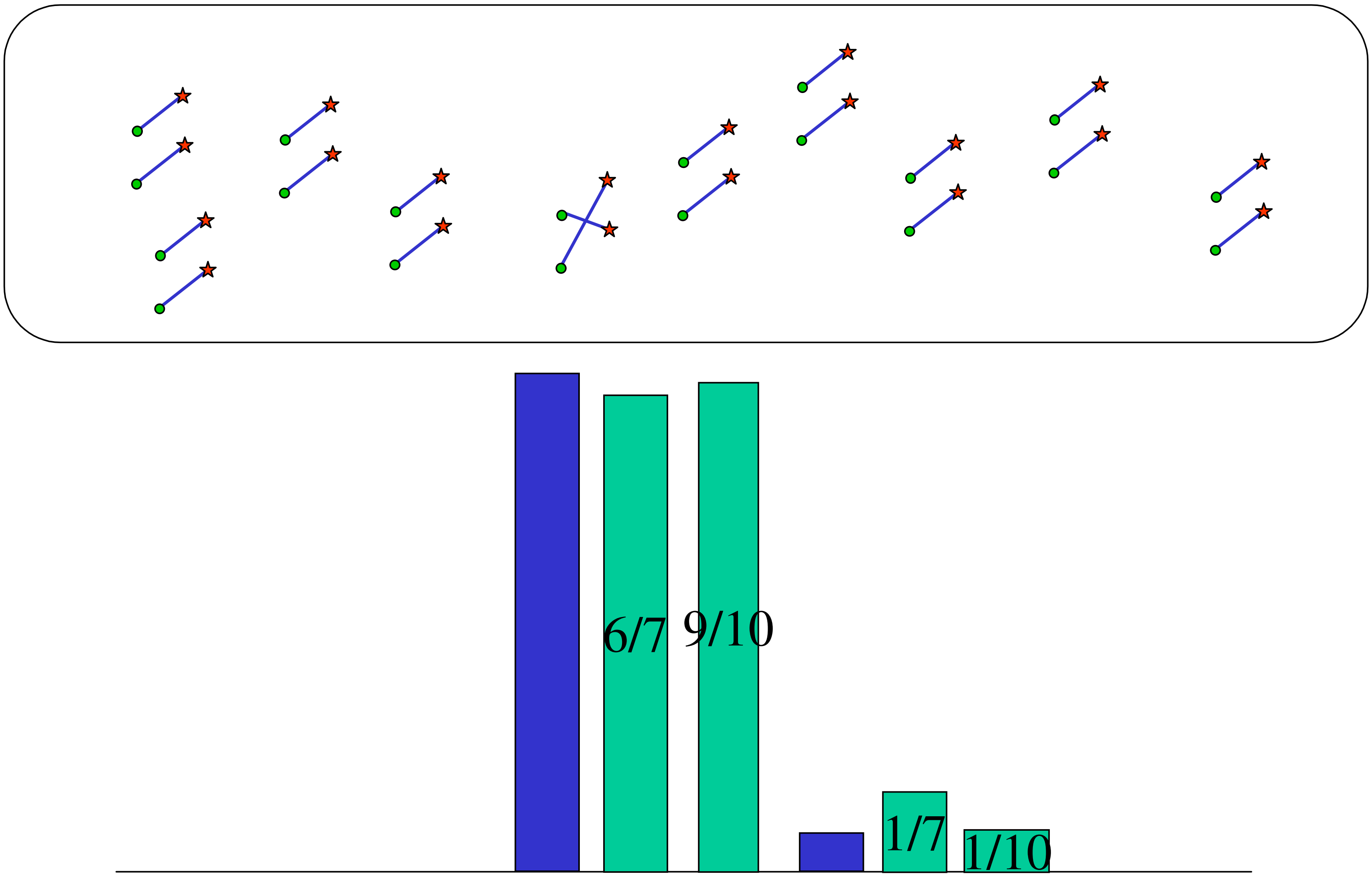
- To get marginals, we need a distribution
- To approximate a distribution, we sample

Monte Carlo Methods

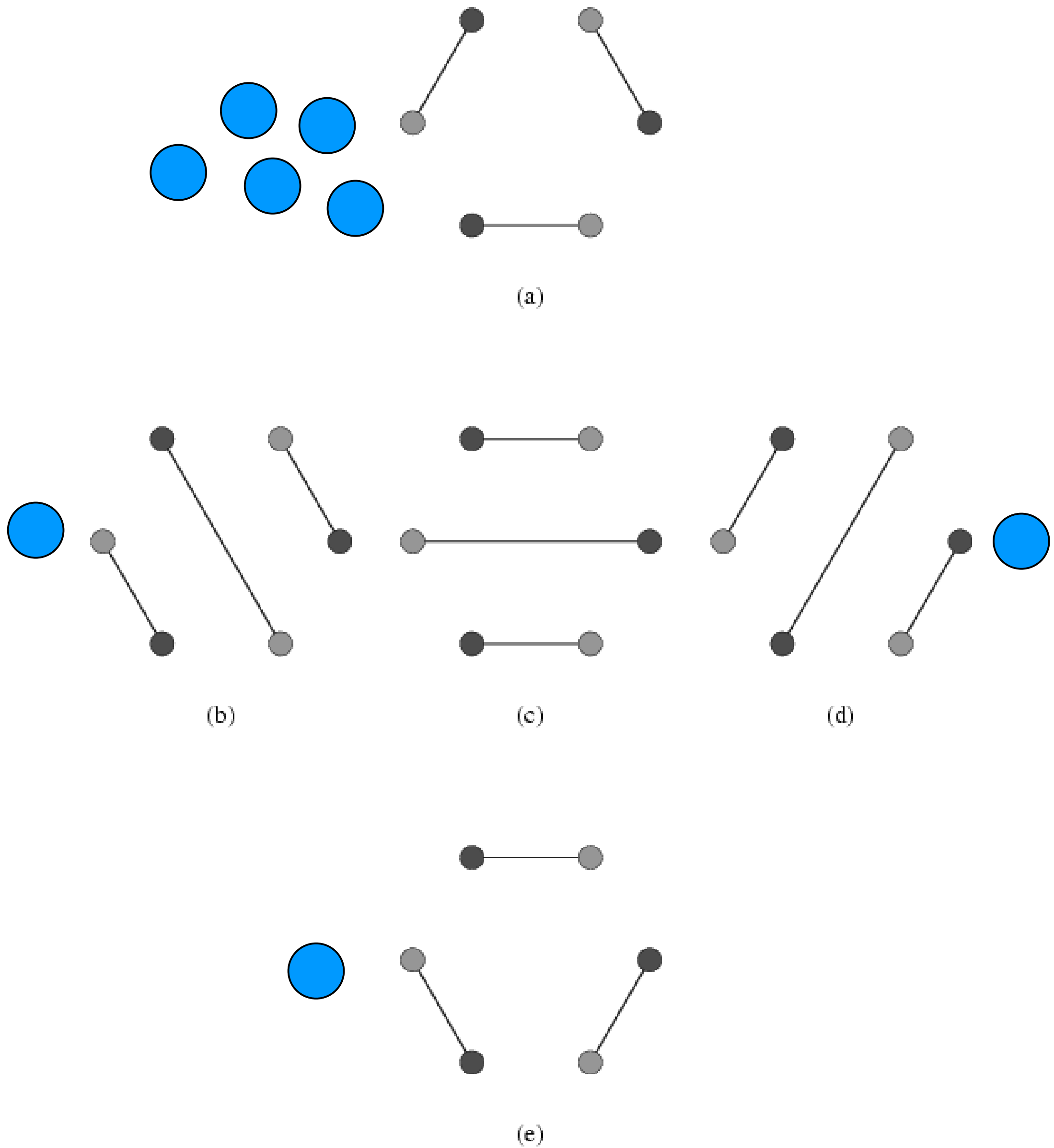
- Samples to approximate a density

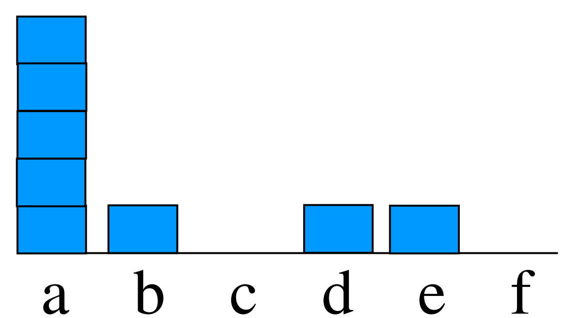


Idea 2: Sample



Markov Chain Monte Carlo





Sampling Assignments using the Metropolis Algorithm (MCMC)

To sample from $P(\mathbf{J}_i|\mathbf{U}_i, \Theta^t)$ we use the Metropolis algorithm:

1. Start with a valid initial assignment $\mathbf{J}_i^0$.

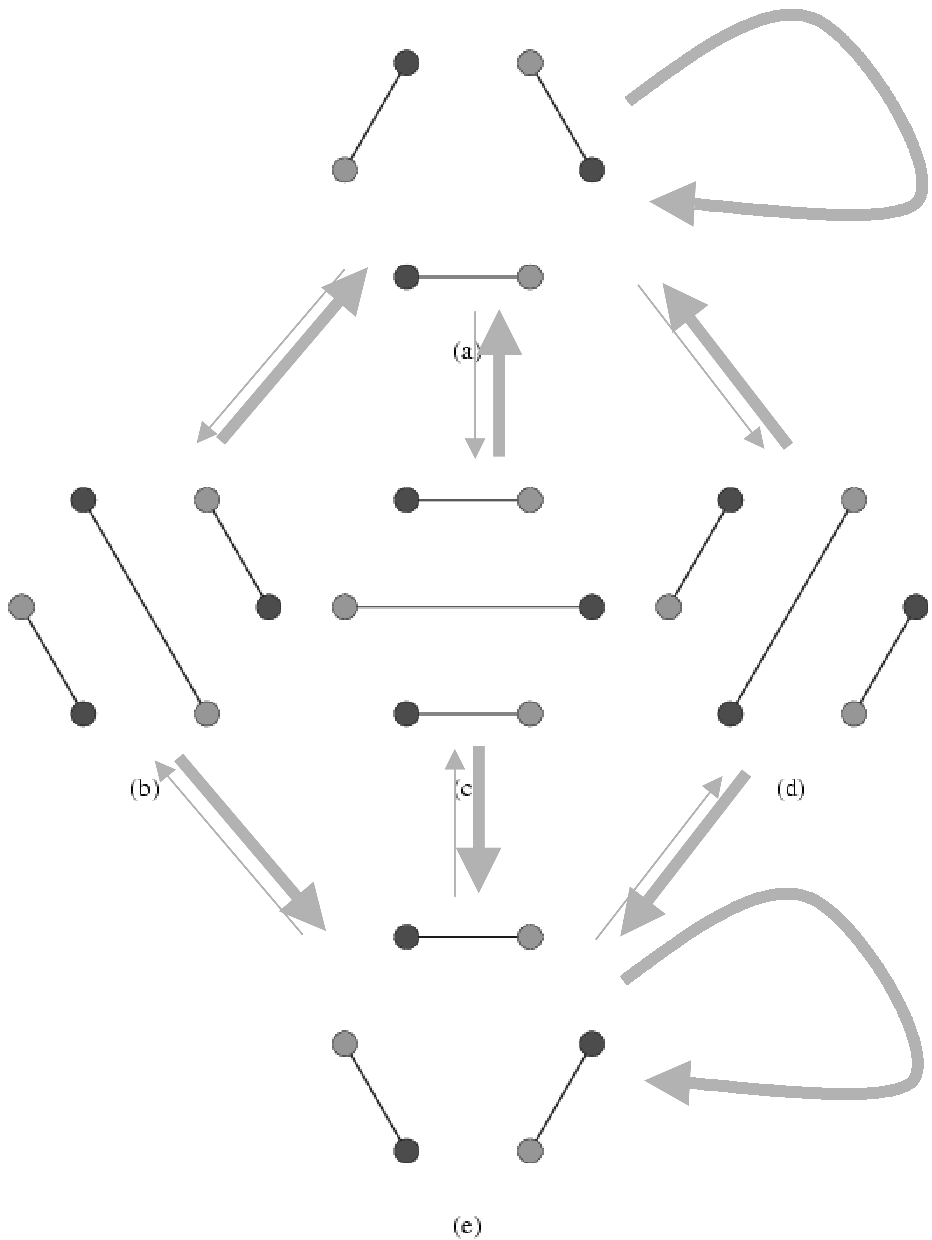
2. Propose a new valid assignment $\mathbf{J}'_i$.

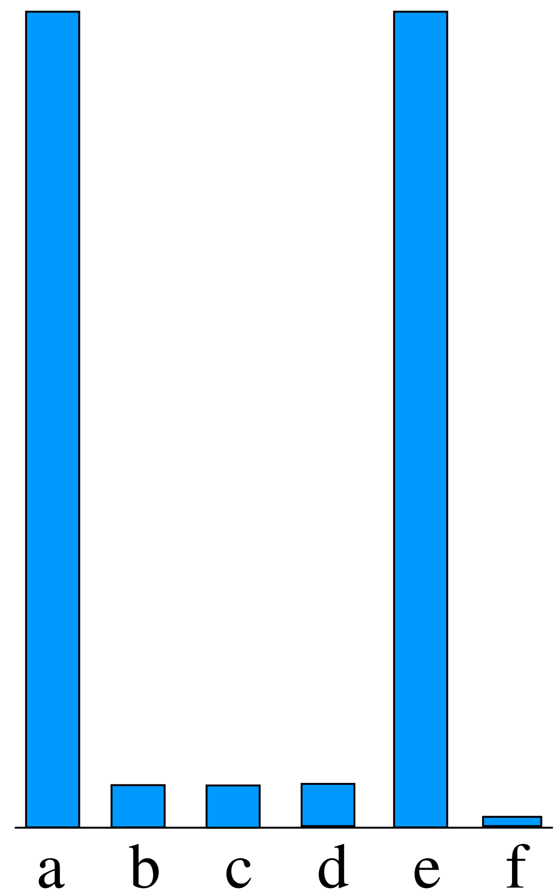
3. Accept with ratio

$$a = \min\left(1,\ \frac{P(\mathbf{J}_i^{'} \mid \mathbf{U}_i, \Theta^t)}{P(\mathbf{J}_i^{r} \mid \mathbf{U}_i, \Theta^t)} \right)$$

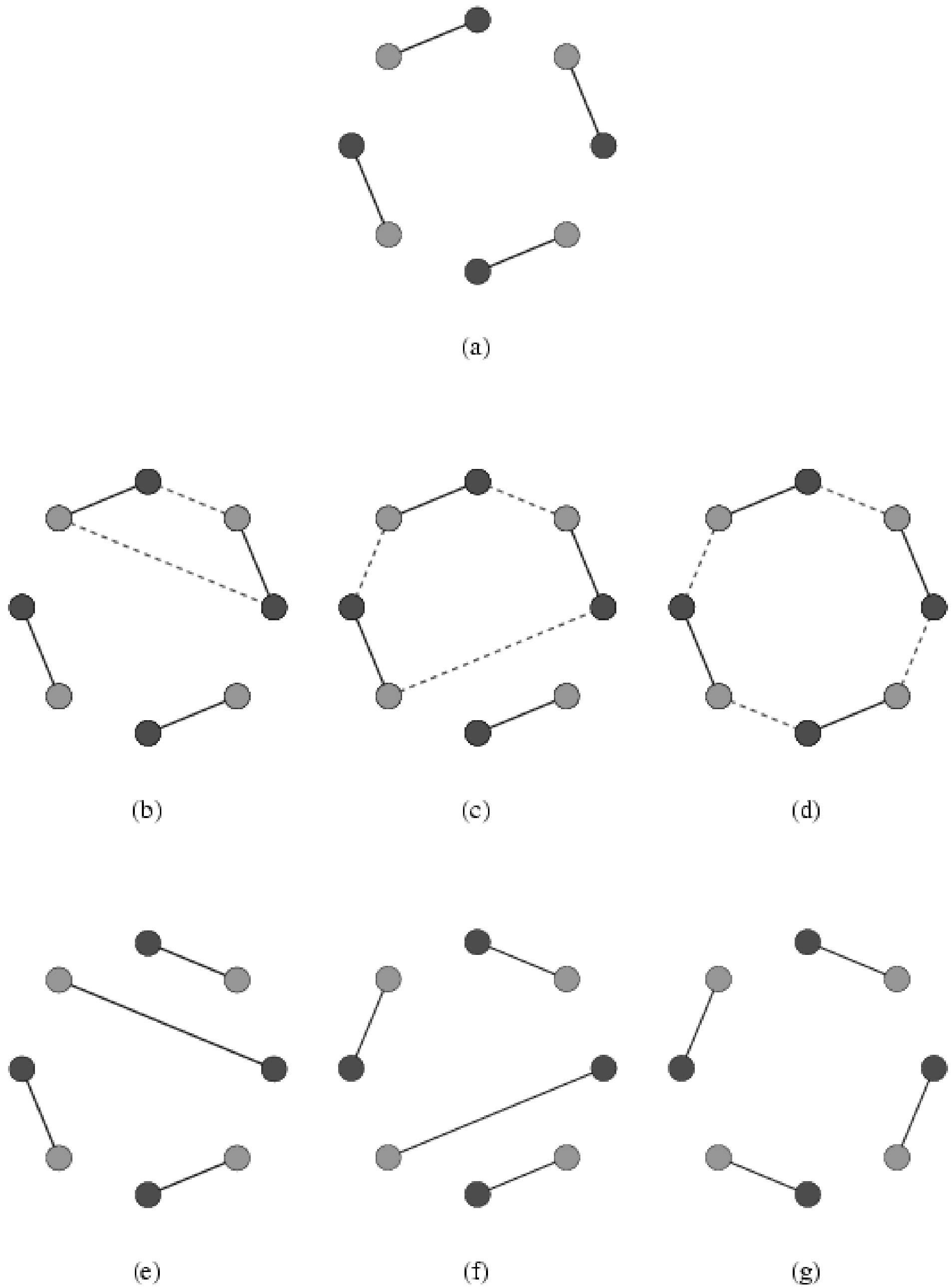
If proposal process is asymmetric: correct **a** (Metropolis Hastings)

Swap Proposals (inefficient)

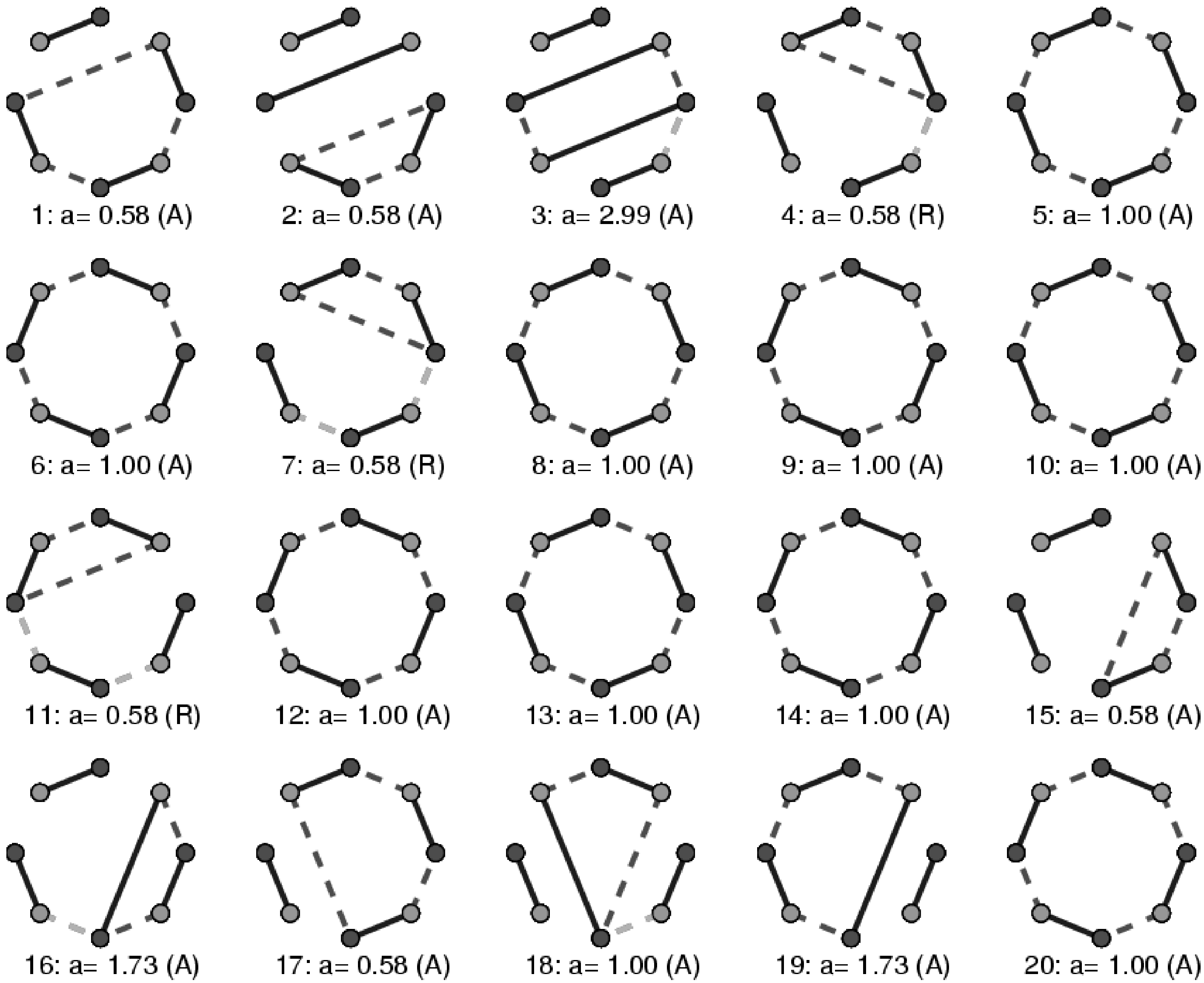




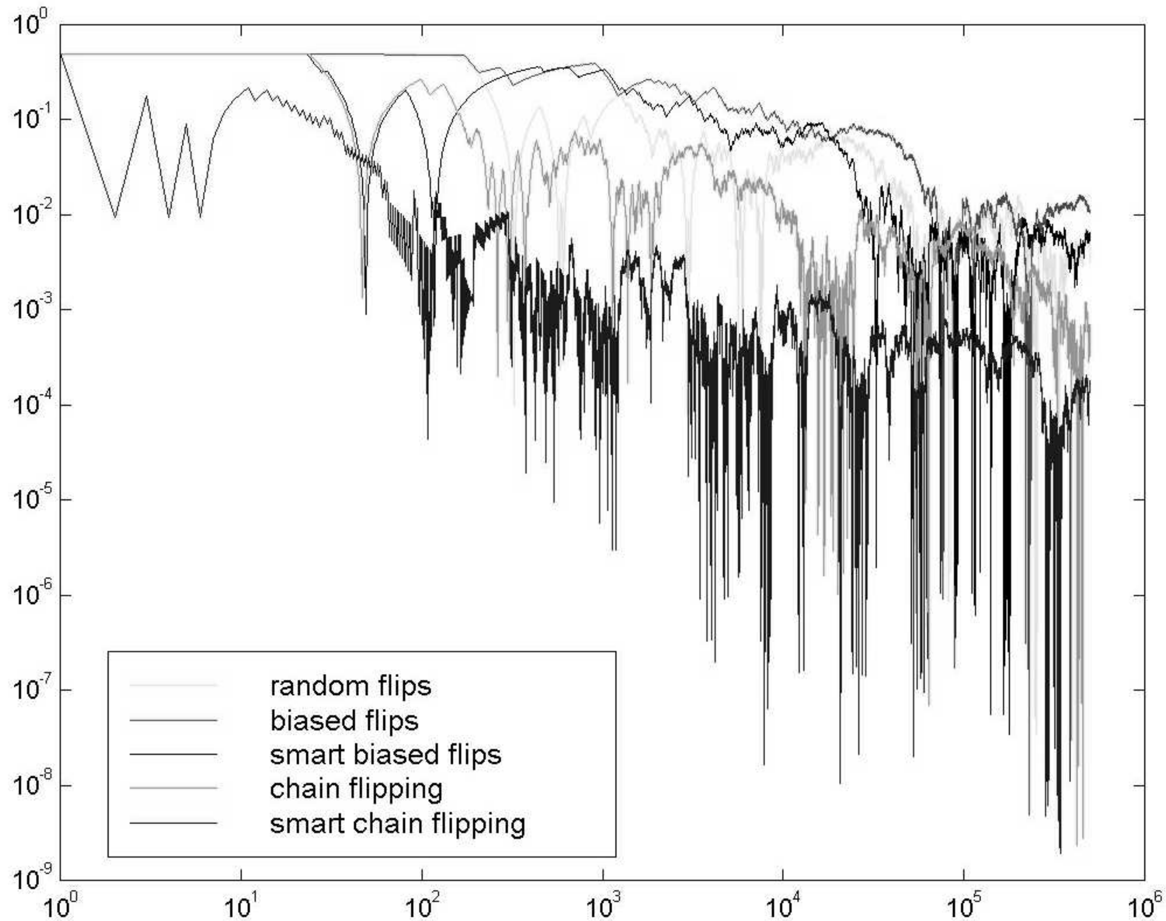
Idea 3: ‘Chain Flipping’ Proposals



Efficient Sampling (Example)



Comparing Proposal Strategies



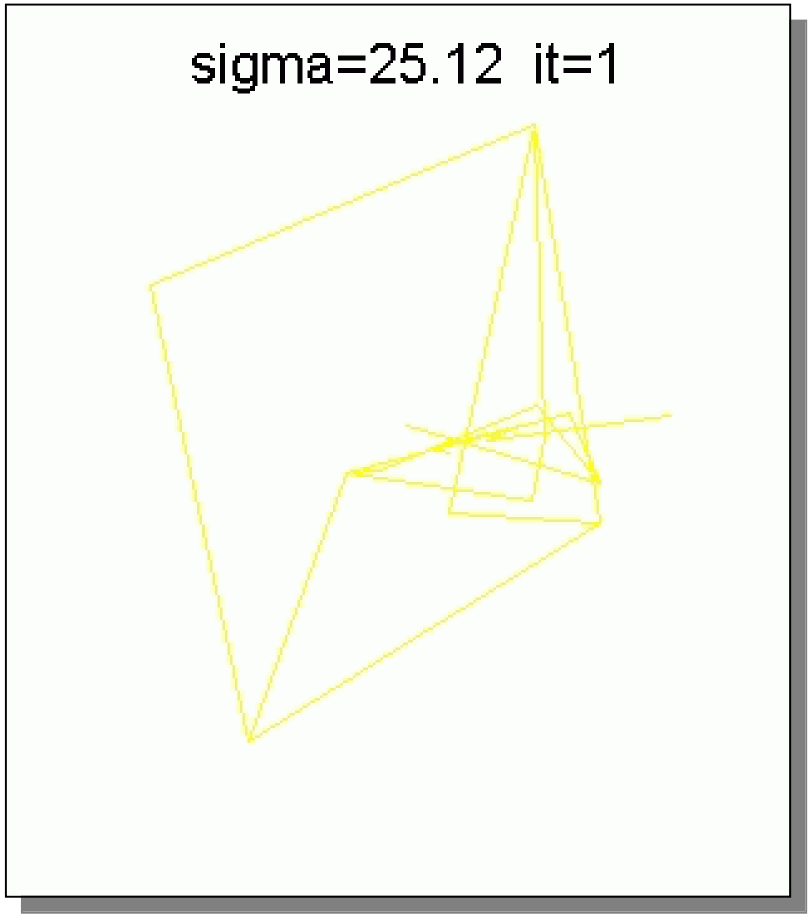
Summary

- Take current estimate for structure/motion
- Project into the image
- Sample assignments using MCMC with Chain Flipping proposal distribution

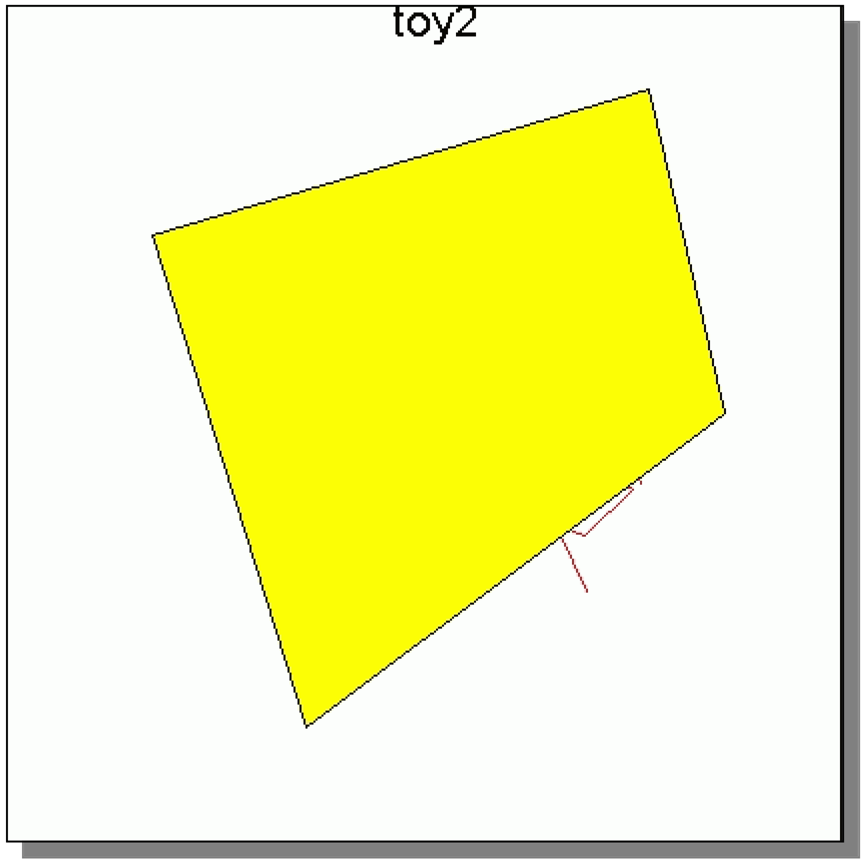
Wire-frame Sequence

Very difficult sequence: very large motion (opposite side)

Wire-frame EM



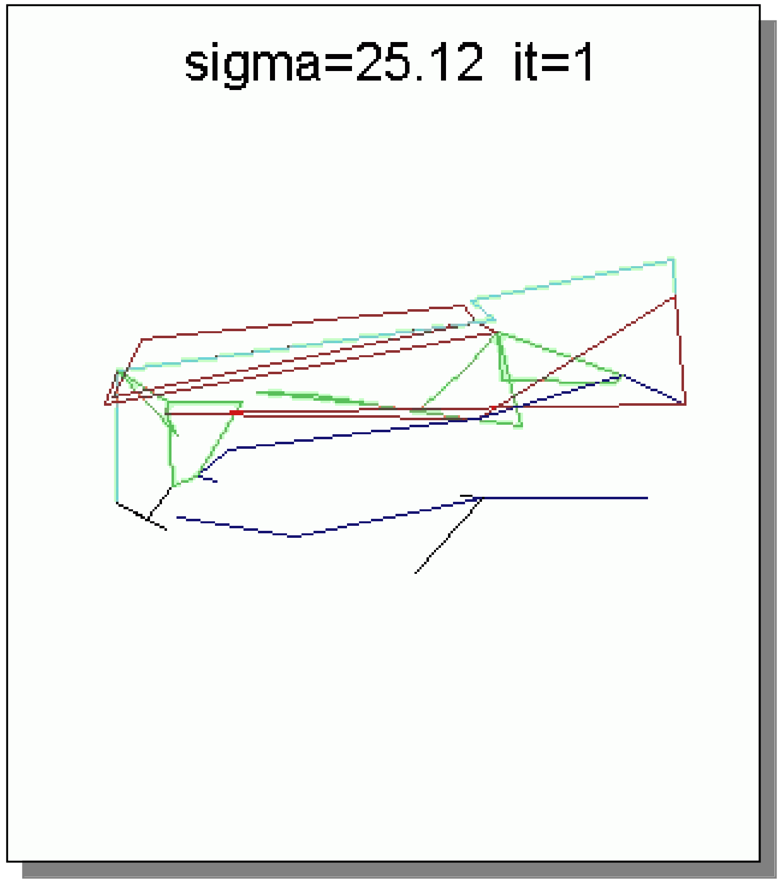
Wire-frame Result



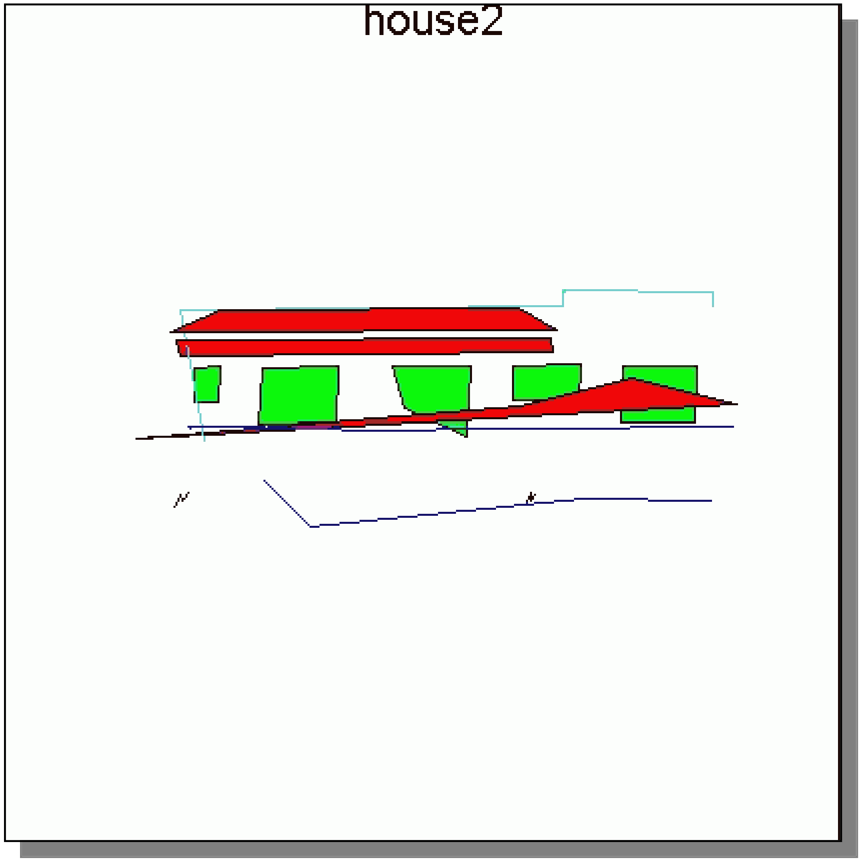
House Sequence with Features

Difficult ! Perspective distortion, large motion, dissimilar features

EM



3D Model



Current Work & Ideas

- Occlusion & Clutter
- Model Selection
  - number of points

- Learning “True Structure” (a la Bayes Net)

Stream Understanding

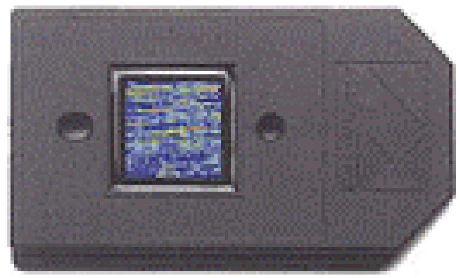
Summary

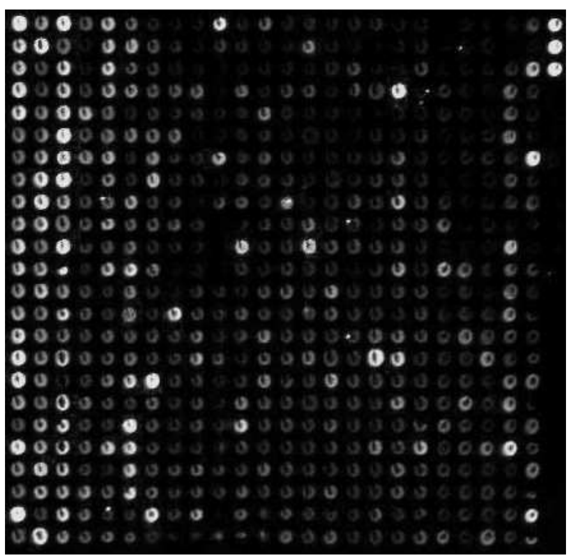
- SFM with unknown correspondences: exponentially harder than regular SFM
  - Step 1: Attack Difficult ML problem using EM
  - Step 2: Create virtual measurements for M step
  - Step 3: Use MCMC algorithm for E step
  - Step 4: Use chain-flipping to attain efficiency
- Good results for non-trivial problems

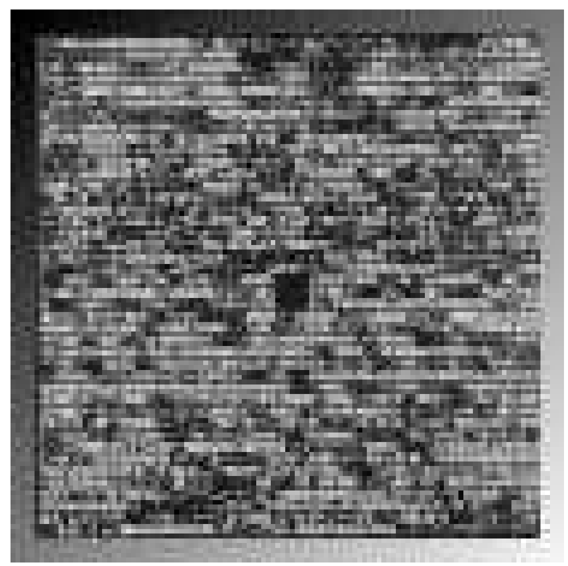
Questions ?

Application in other areas…

Correspondence problems in biology, data mining, medicine,…

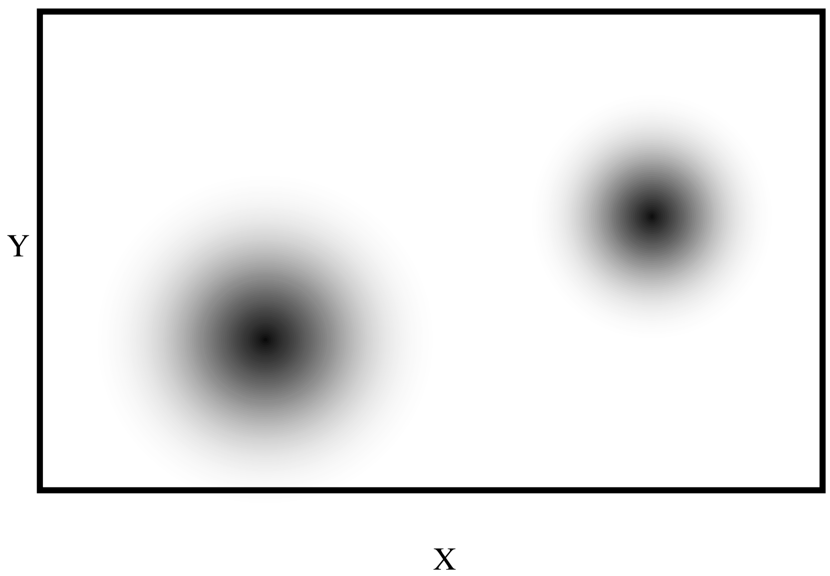


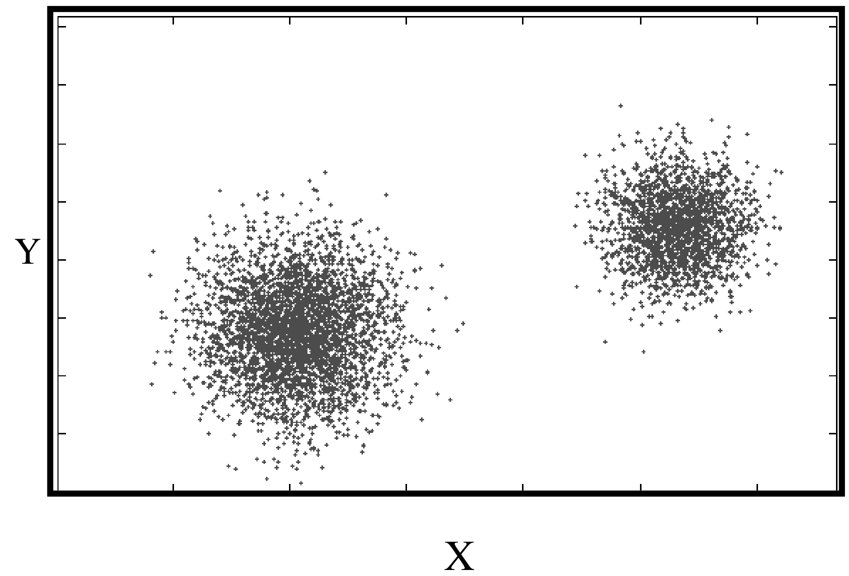




Monte Carlo Localization
(with D. Fox, W. Burgard, S. Thrun)

- Samples to approximate a density





- Monte Carlo Localization: condense2.gif

Notation

Unknowns:

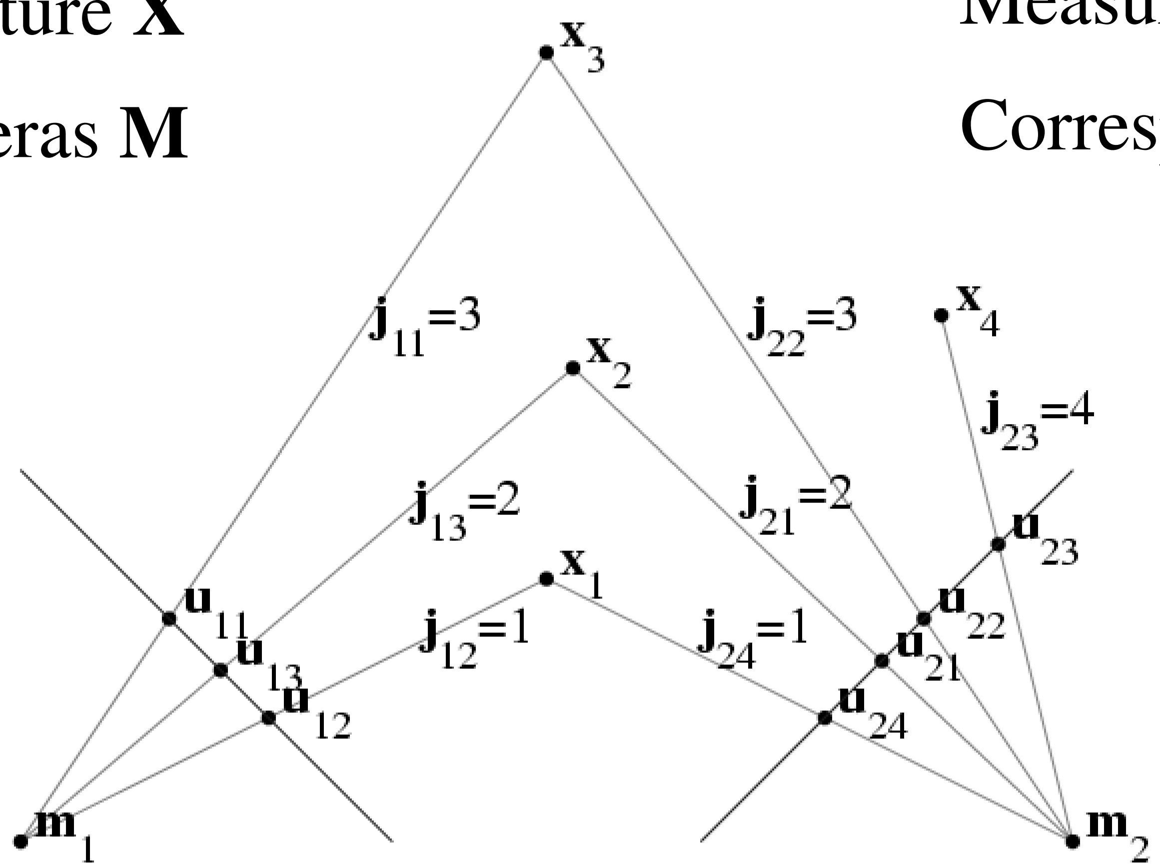
Structure **X**

Cameras **M**

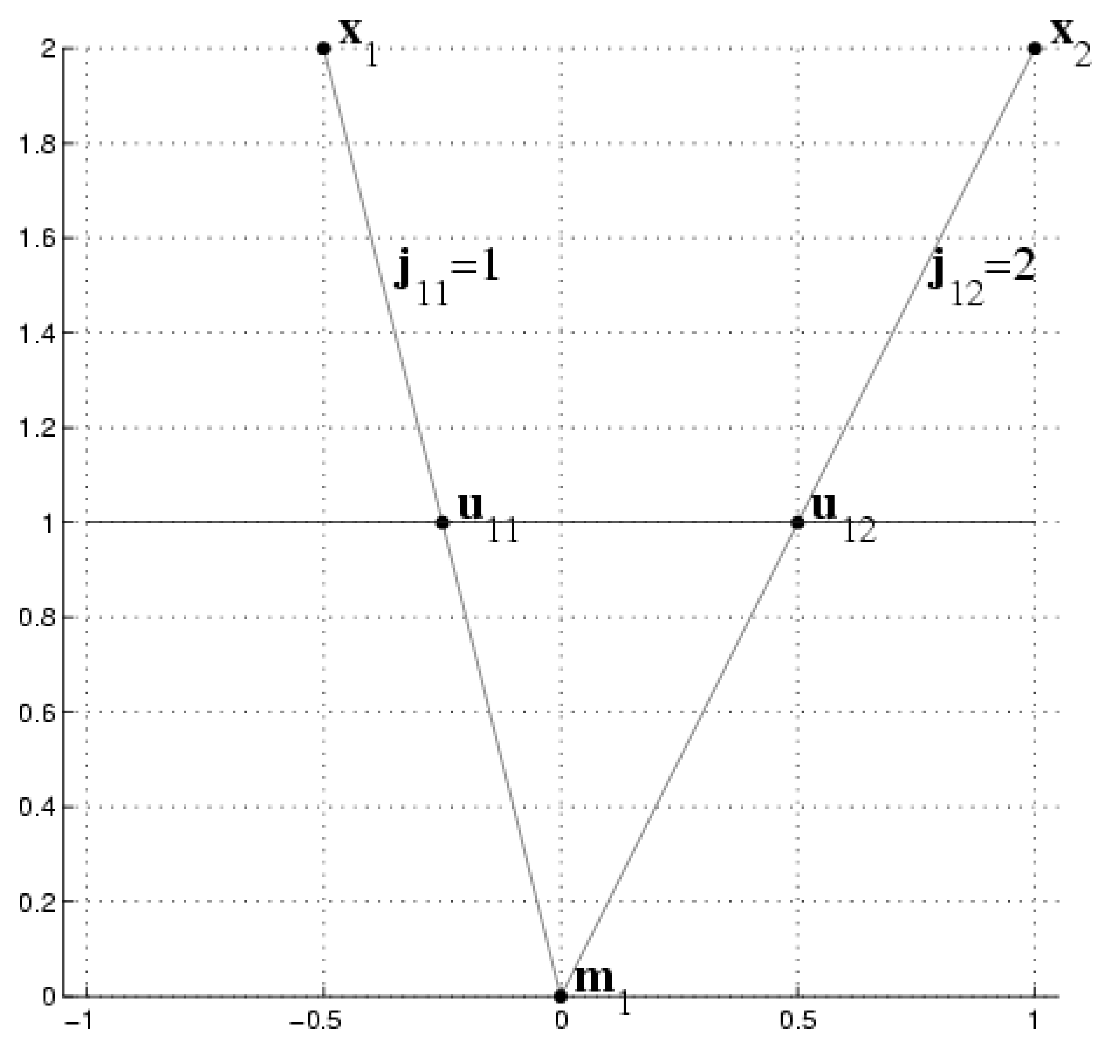
Given:

Measurements **U**

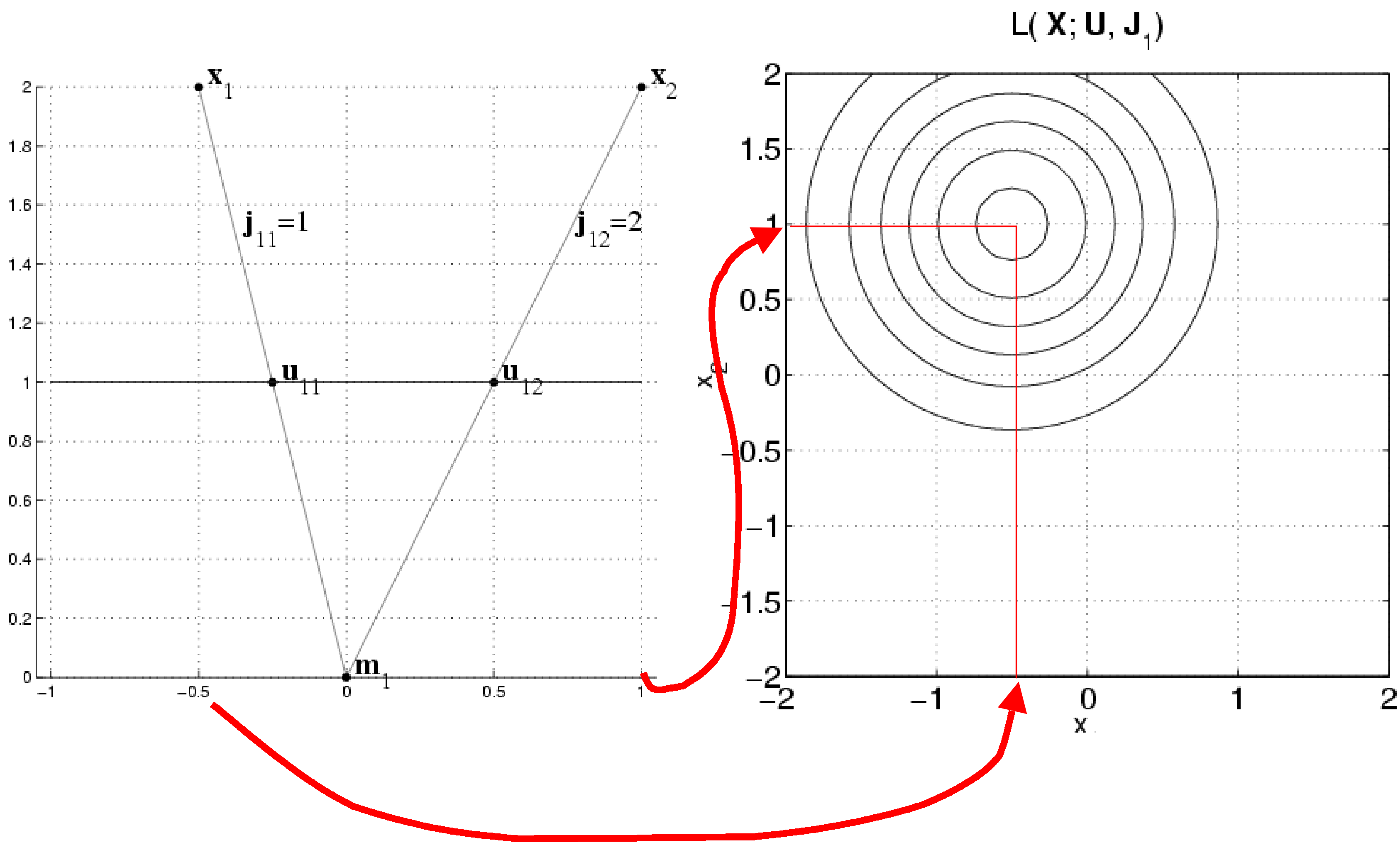
Correspondence **J**



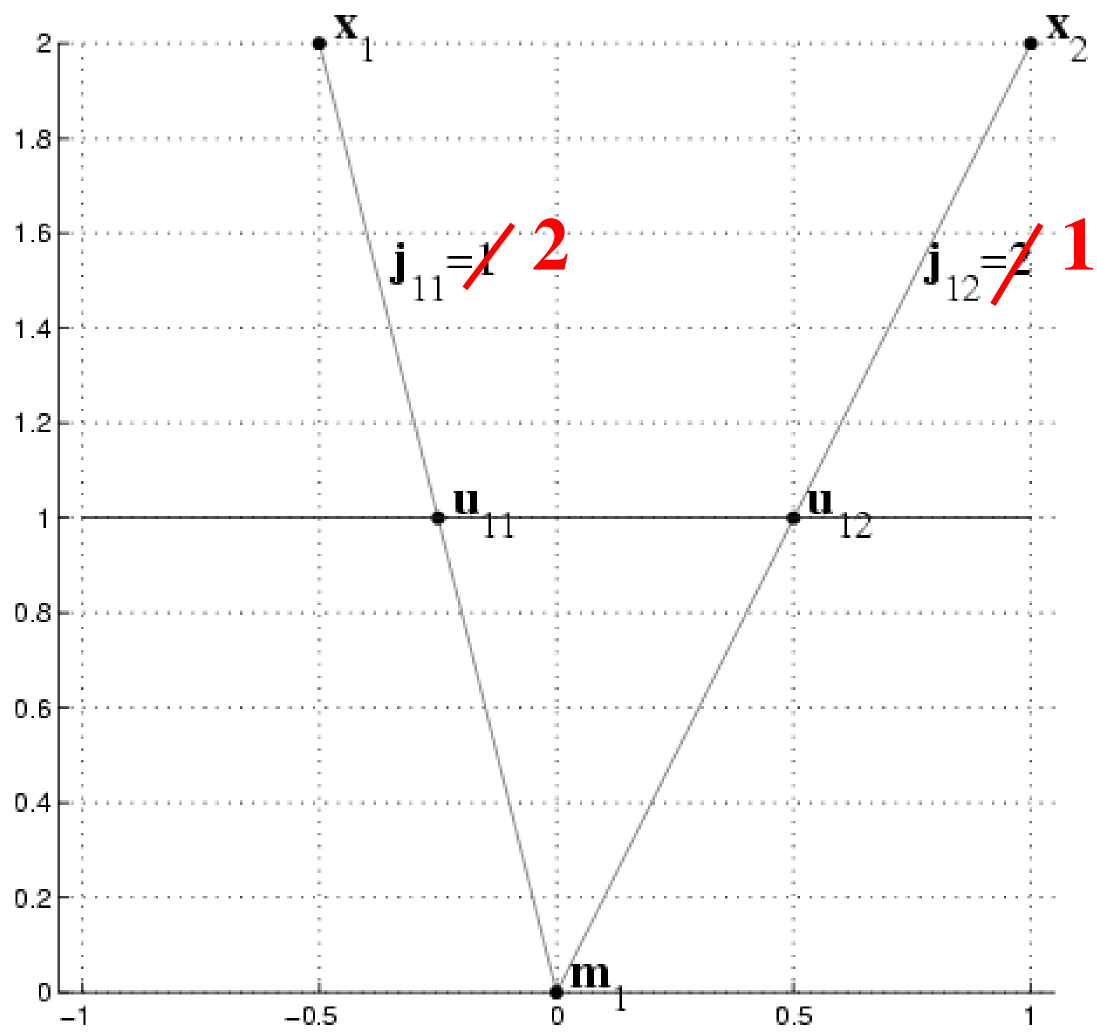
Trivial Example



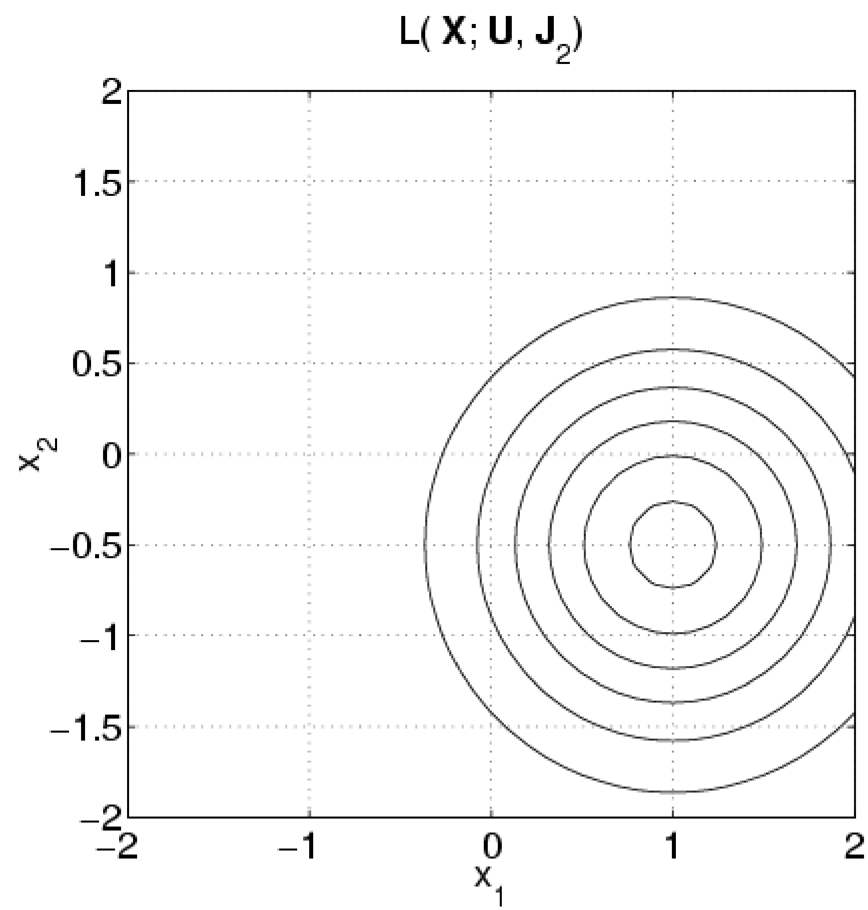
Maximum Likelihood



Alternative Situation



Alternate Universe J_2 !



Monte Carlo EM

- MCEM = Existing Technique
- New = applied to data-association problem
  - Generate samples $J \propto f^t(J)=P(J|U,\Theta^t)$
  - Use Samples To Approximate E-step

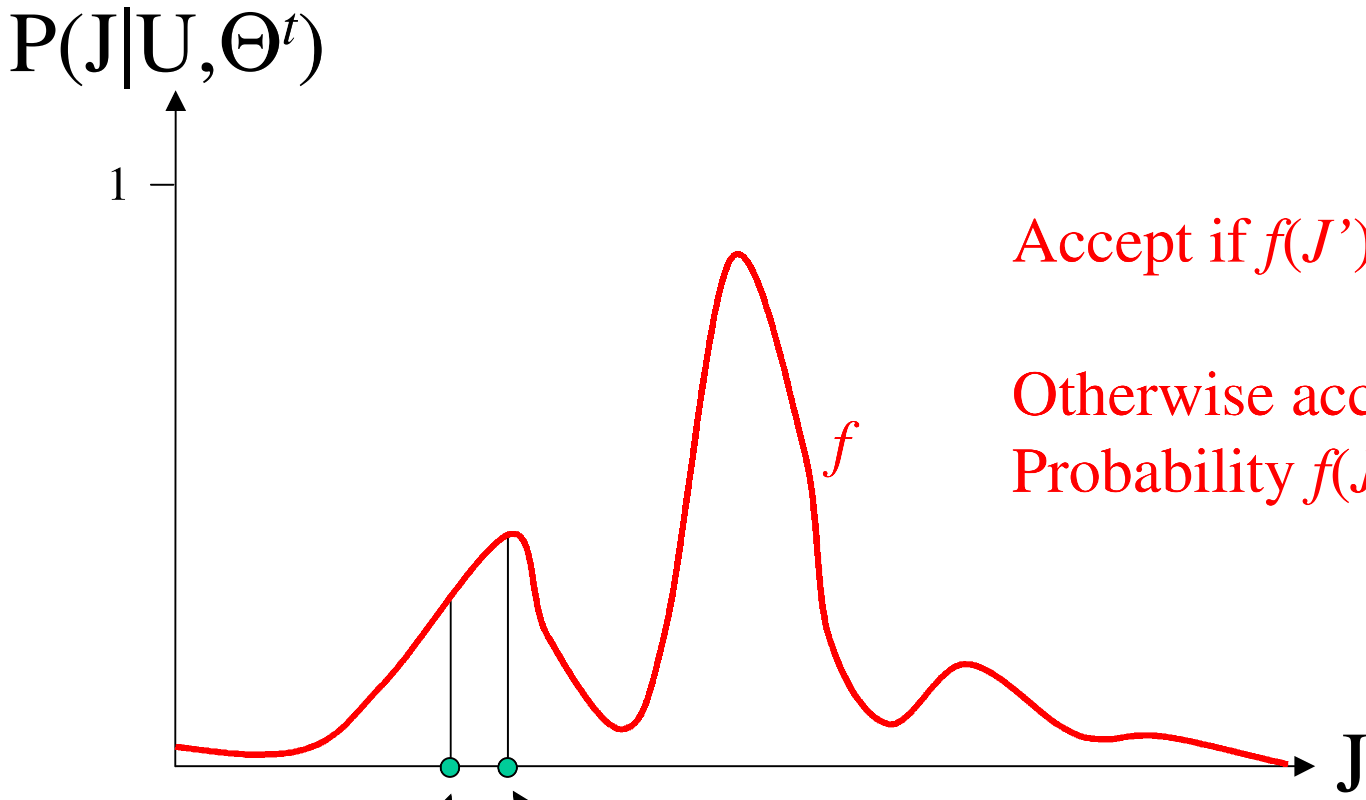
Markov Chain Monte Carlo

- No closed form sampler !
- Construct a Markov Chain over the space of correspondence such that stationary distribution is

$$f^t(J)=P(J|U,\Theta^t)\ )$$

- Works in high-dimensional spaces

Markov Chain Monte Carlo



Accept if *f*(*J'*) > *f*(*J*)

Otherwise accept with Probability *f*(*J'*) / *f*(*J*)

J' Add small random value to J

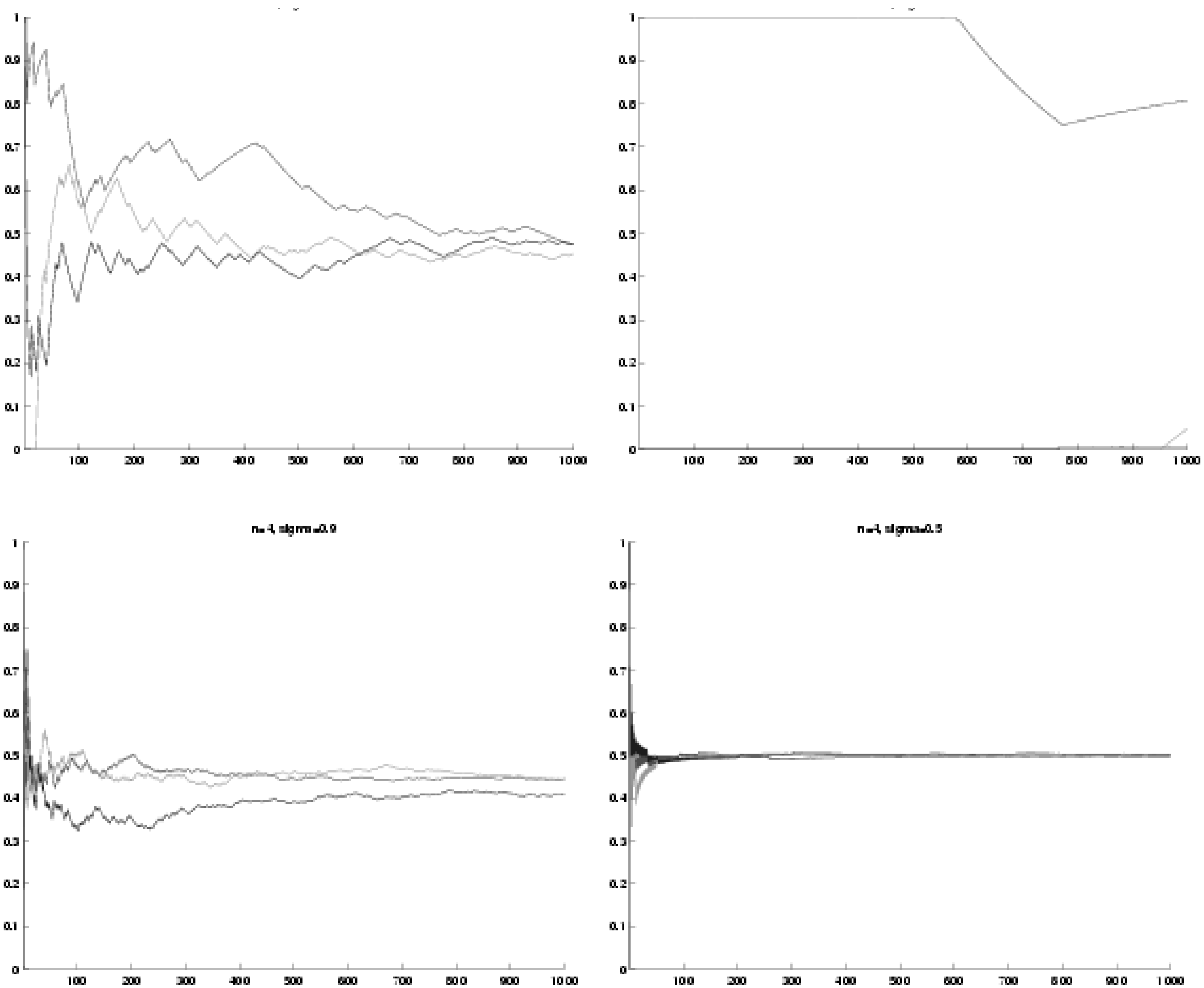
Randomly guess a value for J

Convergence

σ = 0.9 | σ = 0.5

Swap proposal

Chain flipping proposal



Five Images of a House

The Big Picture

A Digital Explosion

- Audio, pictures, video
- Store, retrieve, enhance, process…
- Speech understanding
- Image understanding

My recipe

Probabilistic reasoning algorithms

Incorporate domain knowledge by modeling

Use sound approximations for efficiency

With Dieter Fox:Monte Carlo Robot Localization

Correspondence in General

- Databases: record matching
- Belief Net Structure
- Computational Biology
- Clustering
- …