Mixed Scale Motion Recovery

James Davis

Ph.D. Oral Presentation

Advisor – Pat Hanrahan
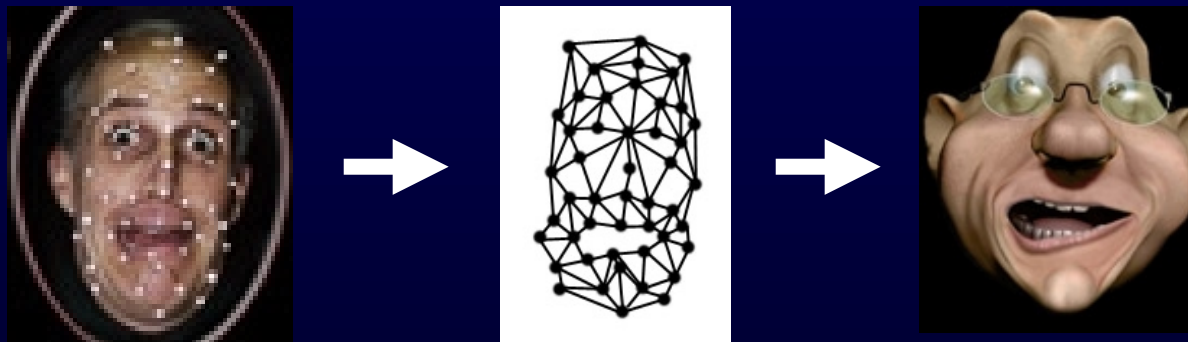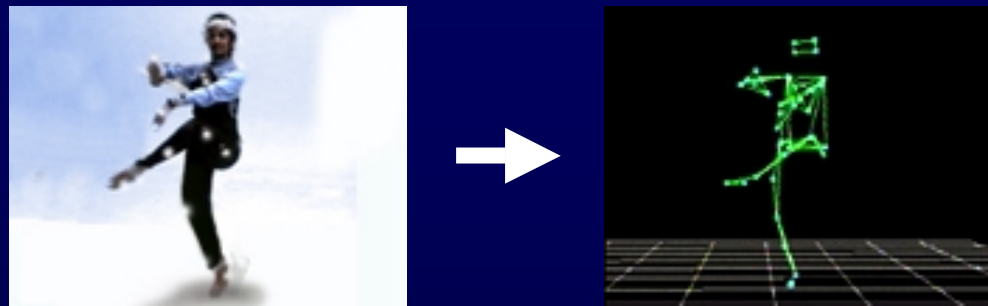
Aug 2001

High level goal

- **Recover motion**
  - Large working volume
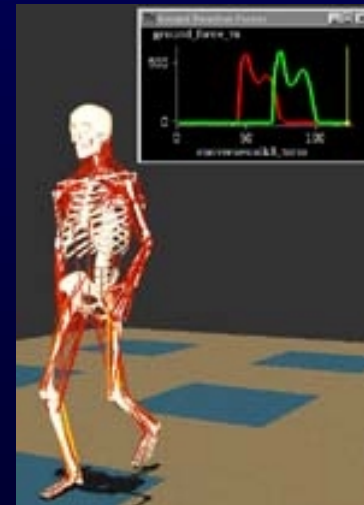 - Extreme detail

Current technology

- **Acquire real motion as computer model**
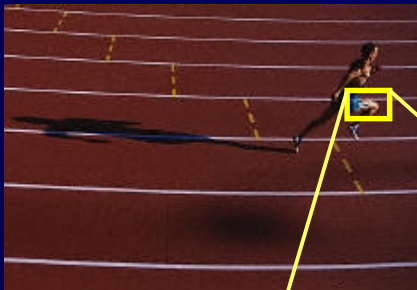

- **Fixed resolution vs. range ratio**

Applications of motion recovery

- **Animation**
- **Athletic analysis**
- **Biomechanics**

Mixed scale domains

- **Detailed motion within a larger volume**

Problem domain characterization

- **Multiple scales of motion**
- **At individual scales**
  - Working volume is local
  - Working volume is moving
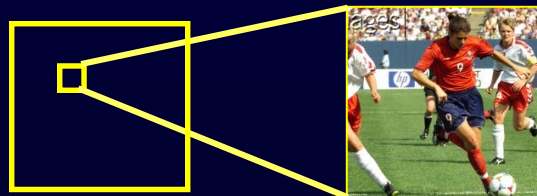
Hierarchical paradigm

- **Explicitly expresses motion hierarchy**
  - Motion recovery drives sub-region selection
  - Sub-region selection defines next scale
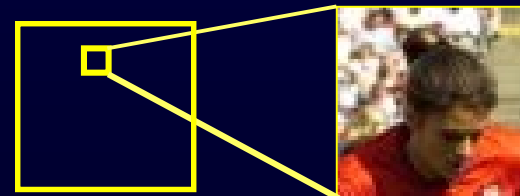- **Multiple designs possible within framework**

| Motion recovery | → | Sub-region selection | → | Motion recovery | → | Sub-region selection | → | Motion recovery |

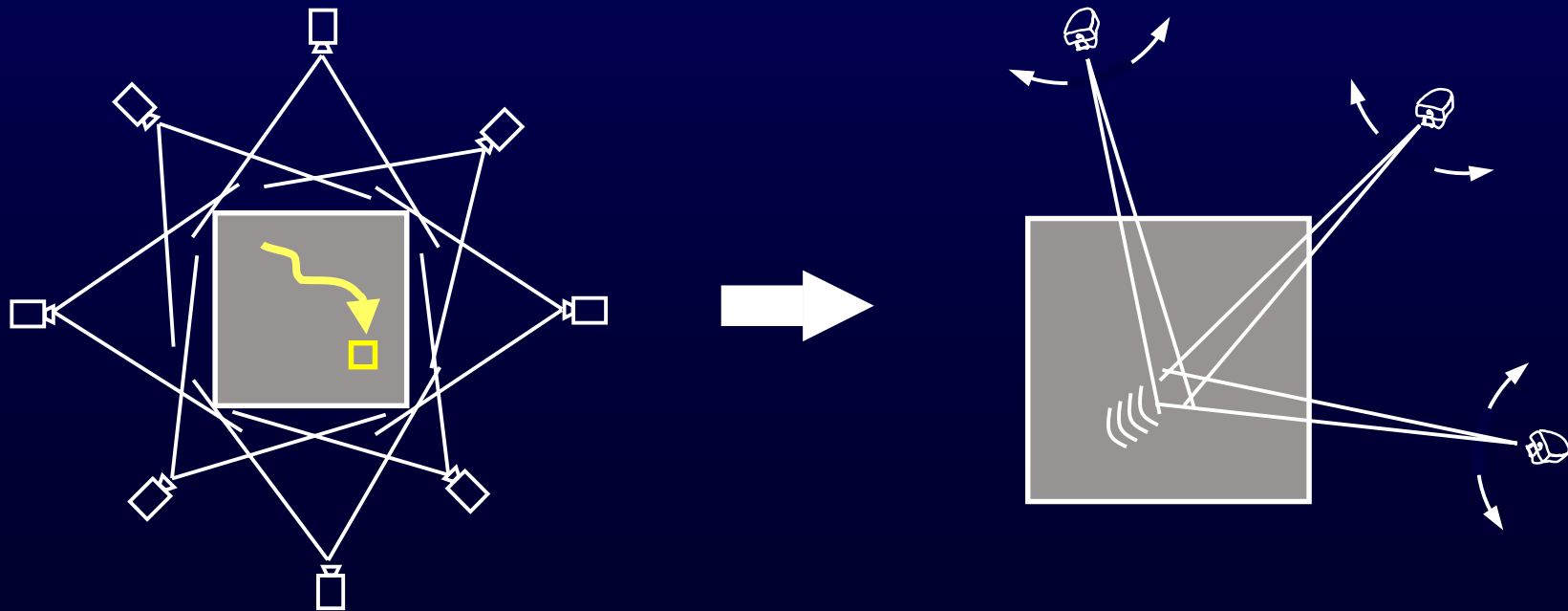**Large scale**   **Medium scale**   **Small scale**
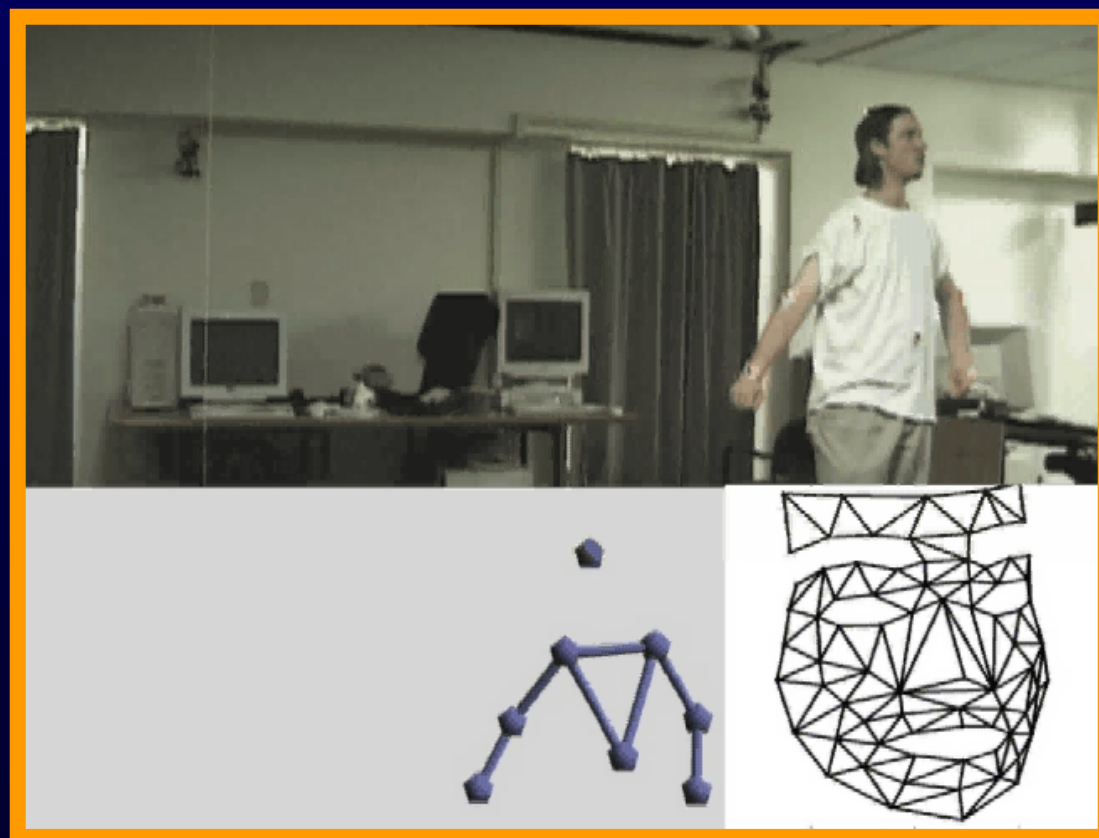
My design

- **Multi-camera large scale recovery**
  - Covers large volume
  - Robust to occlusion
- **Pan-tilt multi-camera small scale recovery**
  - Automated camera control
  - High resolution imaging

Demonstration of my system

- **Body moves on room size scale**
- **Face deforms on much smaller scale**
- **Simultaneous capture**

Desirable system properties

- **High resolution/range ratio**
- **Scalable**
- **Occlusion robustness**
- **Runtime automation**

Related Work

- **Traditional motion recovery**
  - [Vicon] [MotionAnalysis] [Guenter98]

- **Simple pan/tilt systems**
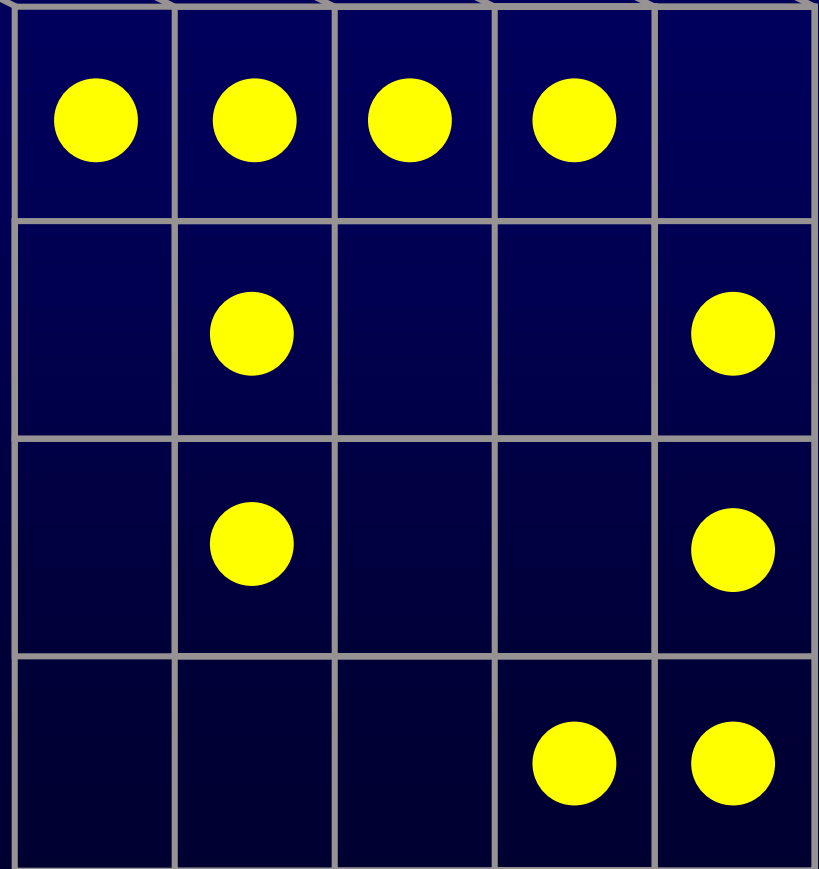 - [Sony EVI-D30] [Fry00]

- **2D guided pan/tilt systems**
  - [Darrell96] [Greiffenhagen00]

- **Human controlled cameras**
  - [Kanade00]

|	Recovers motion	Runtime automation	Occlusion robustness	Scalable	High resolution/range ratio
Traditional motion recovery	●	●	●	●	
Simple pan/tilt systems		●			●
2D guided pan/tilt systems		●			●
Human controlled cameras				●	●

Contributions

- **Framework for mixed scale motion recovery**
  - Hierarchical paradigm
  - Data driven analysis
  - Model based solutions

- **Specific system design**
  - High resolution/range ratio
  - Scalable
  - Robust to occlusion
  - Automated

- **Application to simultaneous face-body capture**

Talk outline

- **Introduction**
- **Framework**
  - Hierarchical paradigm
  - Data driven analysis
  - Model based solutions
- **System implementation**
  - Large scale recovery
  - Sub-region selection
  - Small scale recovery
- **End-to-end video**
- **Summary and future work**
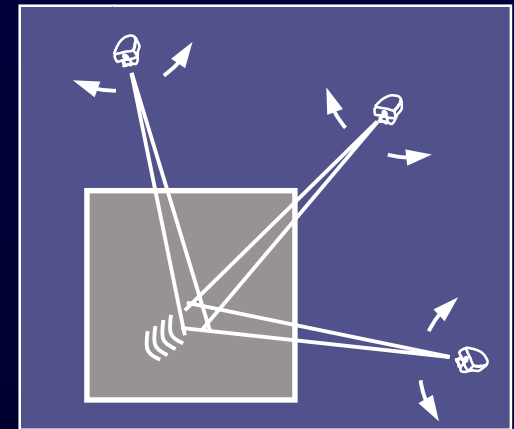
Relation of system to hierarchy

Large scale
motion recovery
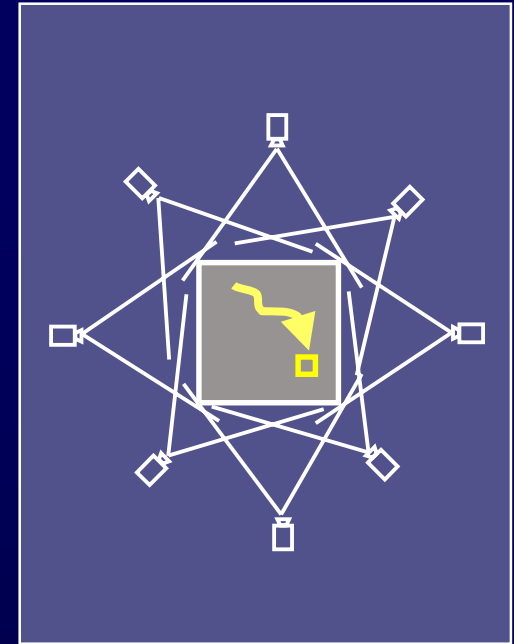
Sub-region selection
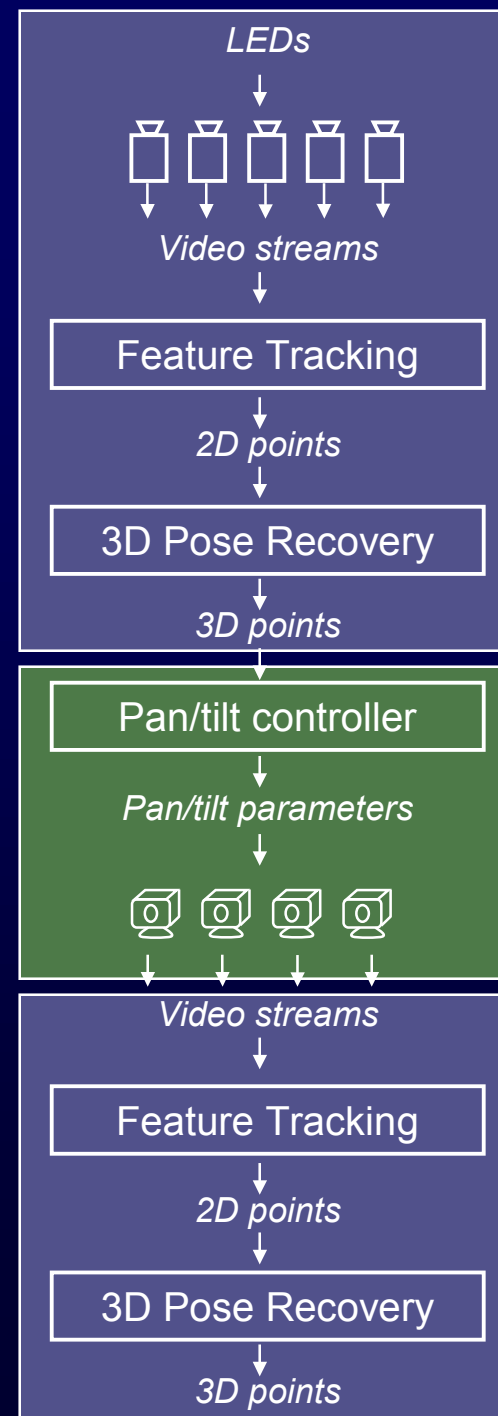
Small scale
motion recovery

System overview

Large scale
motion recovery

Sub-region selection
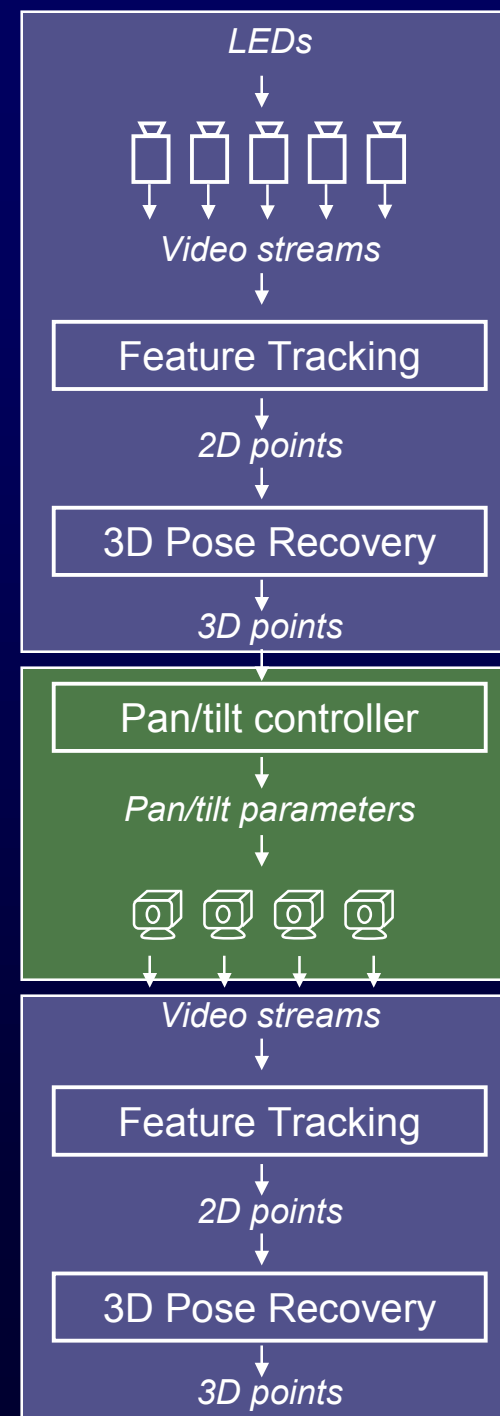
Small scale
motion recovery

LEDs

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Pan/tilt controller

Pan/tilt parameters

Video streams

Feature Tracking

2D points

3D Pose Recovery

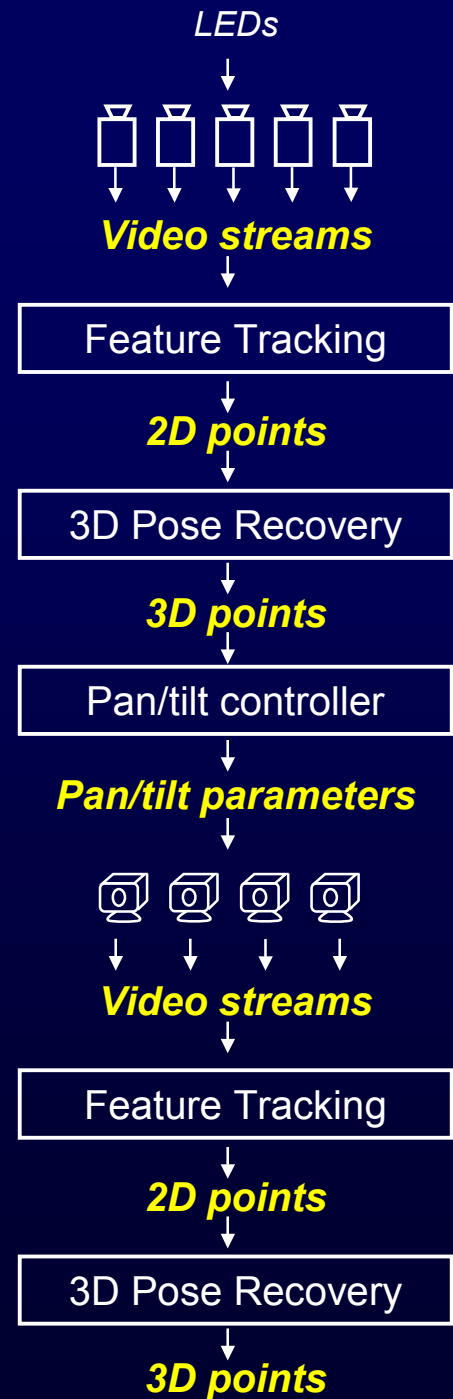*3D points*

Interface is the challenge

- **Large/small scale similar**
- **Interface requirements differ**

Interfaces

- **Interface critical in end-to-end system**
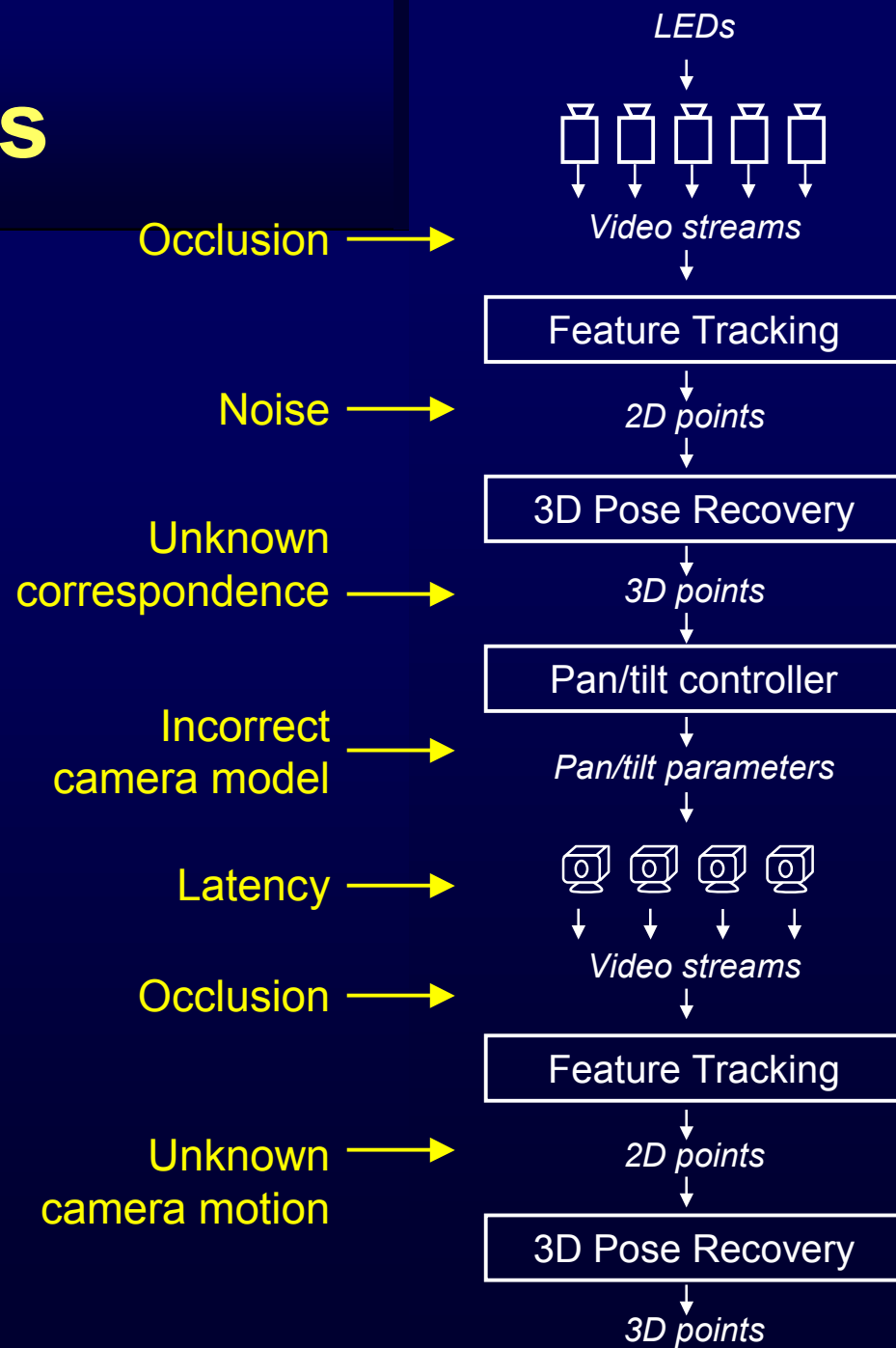- **Often ignored in individual components**

LEDs

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Pan/tilt controller

Pan/tilt parameters

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Data flow

- **System viewed as data flow**
- **Clean interface (data) desirable**

LEDs

Video streams

| Feature Tracking |

2D points

| 3D Pose Recovery |

3D points

| Pan/tilt controller |

Pan/tilt parameters

Video streams

| Feature Tracking |

2D points

| 3D Pose Recovery |

3D points

System challenges

LEDs

↓

⬓ ⬓ ⬓ ⬓ ⬓

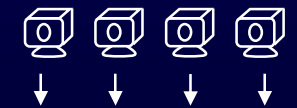↓ ↓ ↓ ↓ ↓

Occlusion →

Video streams

↓

| Feature Tracking |

Noise →

2D points

↓

| 3D Pose Recovery |

Unknown
correspondence →

3D points

↓

| Pan/tilt controller |

Incorrect
camera model →

Pan/tilt parameters

↓

Latency →

▱ ▱ ▱ ▱

↓ ↓ ↓ ↓

Video streams

Occlusion →

↓

| Feature Tracking |

Unknown
camera motion →

2D points

↓

| 3D Pose Recovery |

3D points

Model improves data

LEDs

Video streams

Occlusion →

Feature Tracking

2D points

Model → Noise →

3D Pose Recovery

3D points

Model → Unknown correspondence →

Pan/tilt controller

Pan/tilt parameters

Model → Incorrect camera model →

Model → Latency →

Video streams

Model → Occlusion →

Feature Tracking

2D points

Model → Unknown camera motion →

3D Pose Recovery

3D points

Model improves data

LEDs

Occlusion →

Video streams

| Feature Tracking |

2D points

Noise →

| 3D Pose Recovery |

3D points

Kalman filter →

Unknown correspondence →

| Pan/tilt controller |

Pan/tilt parameters

P/T camera model →

Incorrect camera model →

Prediction →

Latency →

Video streams

Face model →

Occlusion →

| Feature Tracking |

2D points

Unknown camera motion →

| 3D Pose Recovery |

3D points

# Talk outline

- **Introduction**

- **Framework**

  - Hierarchical paradigm

  - Data driven analysis

  - Model based solutions

- **System implementation**

  - Large scale recovery

  - Sub-region selection

  - Small scale recovery

- **End-to-end video**

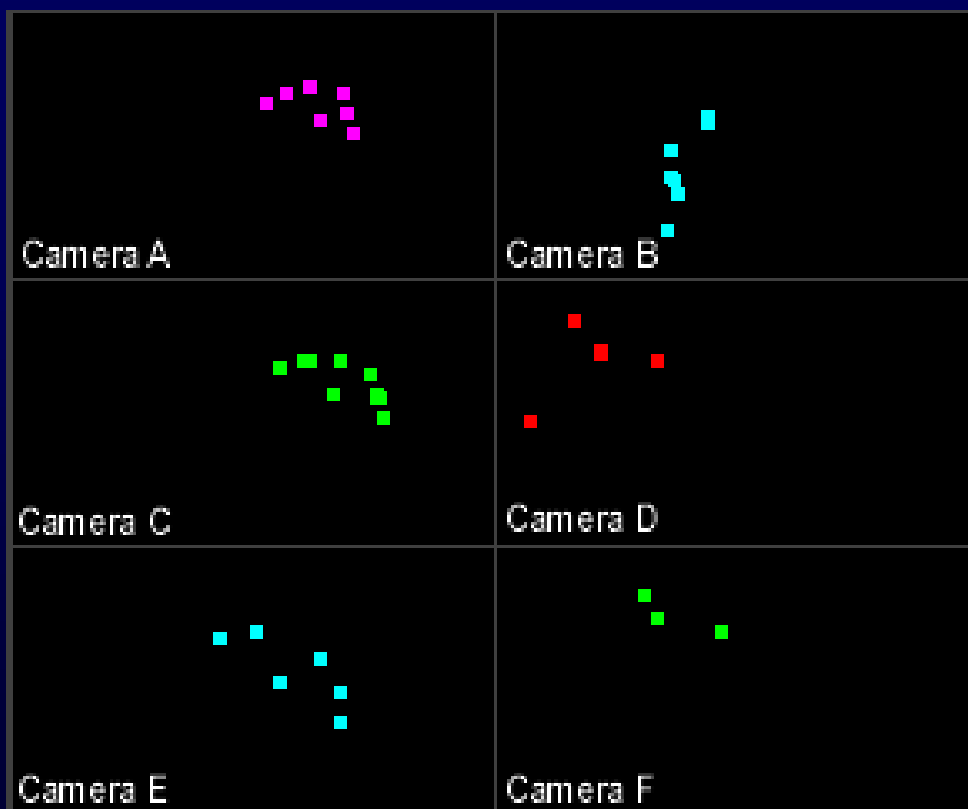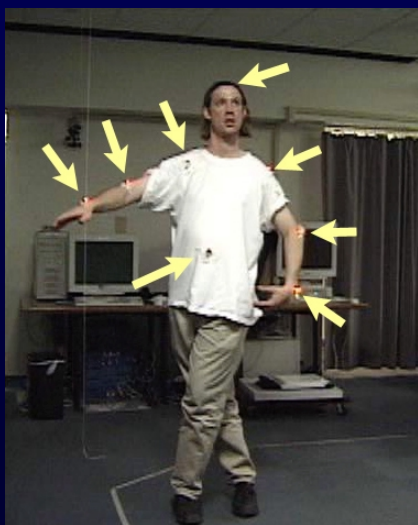- **Summary and future work**

Large scale system
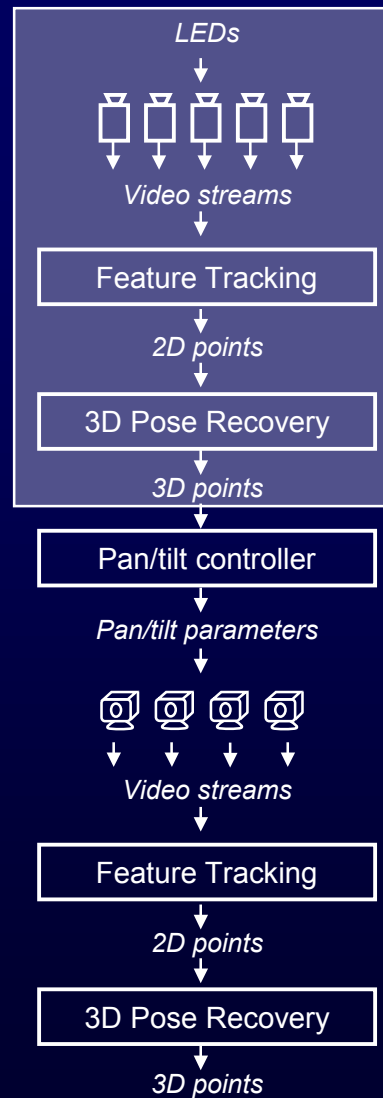
Large scale physical arrangement
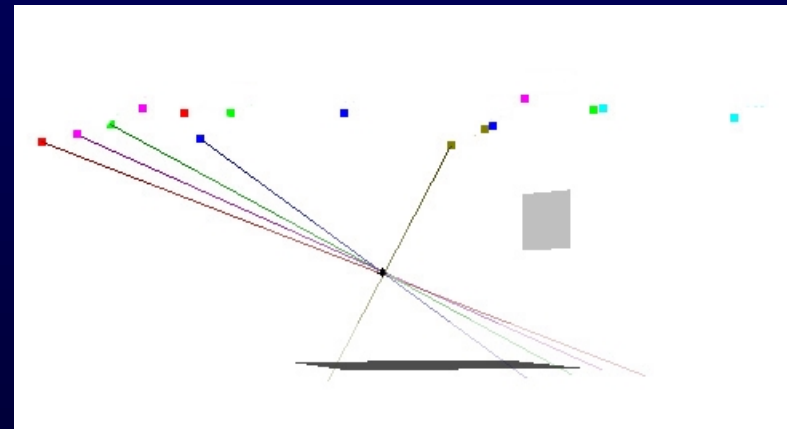
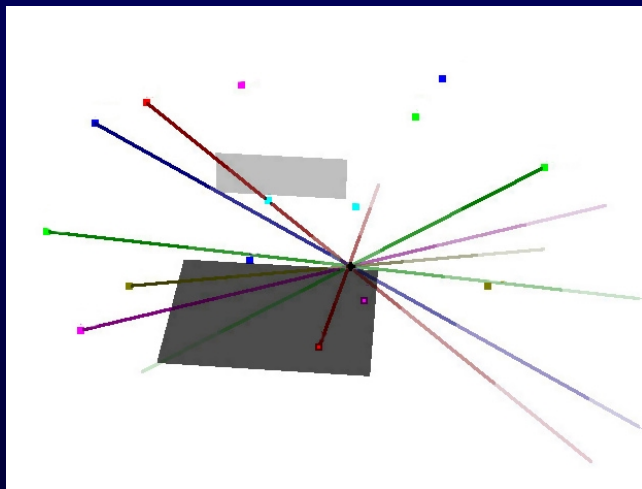- **18 NTSC cameras**
- **18 digitizing Indys**

Large scale features

Camera A

Camera B
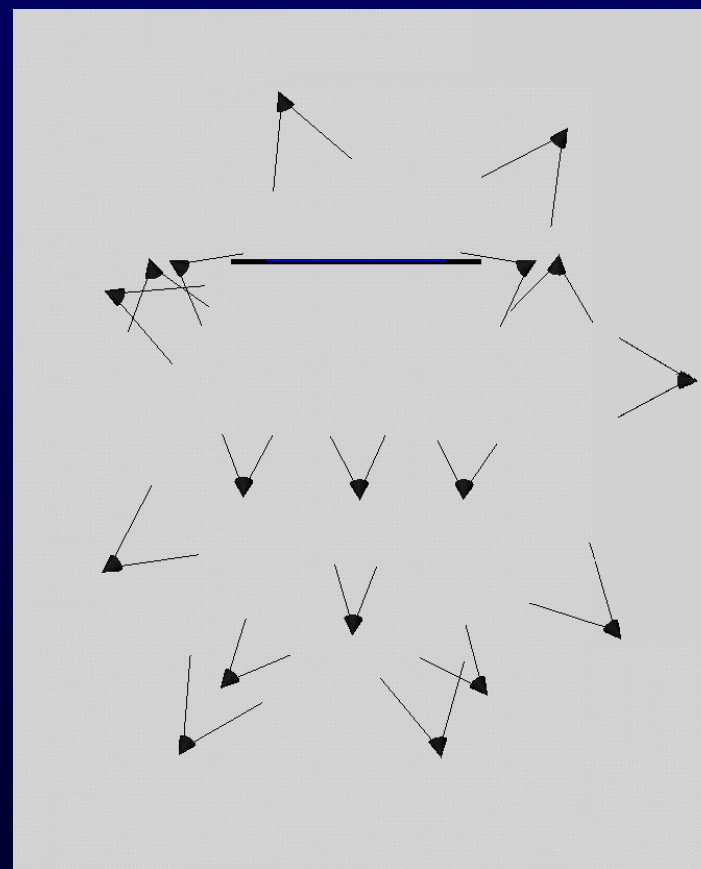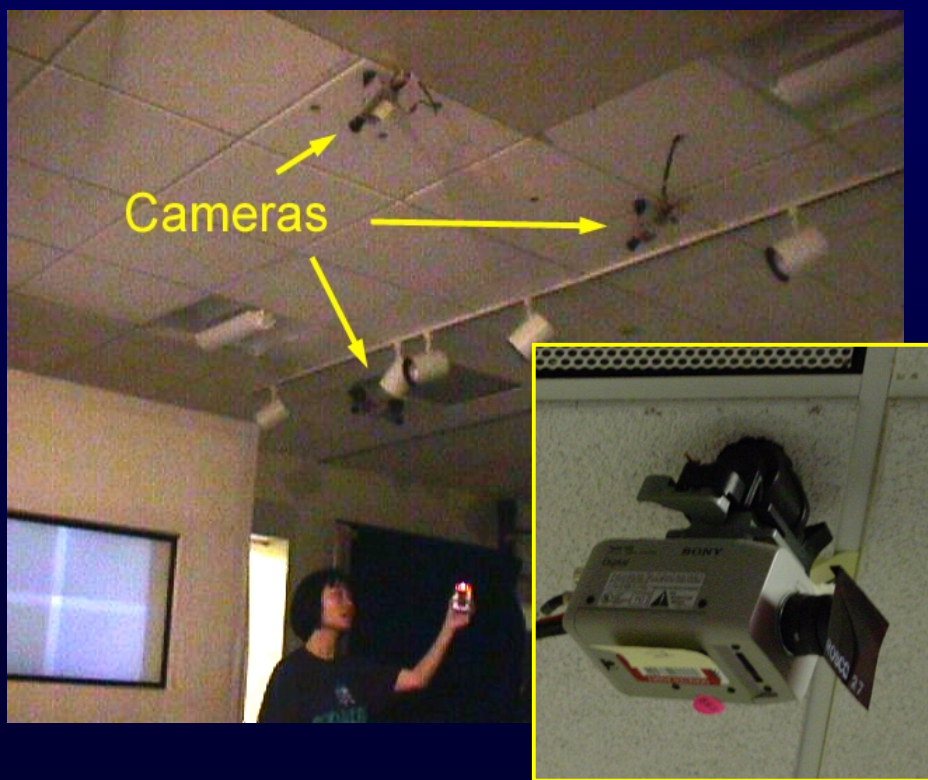
Camera C
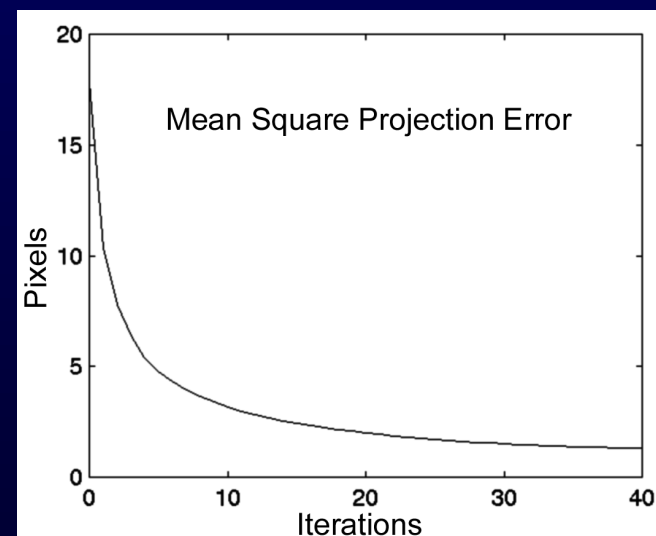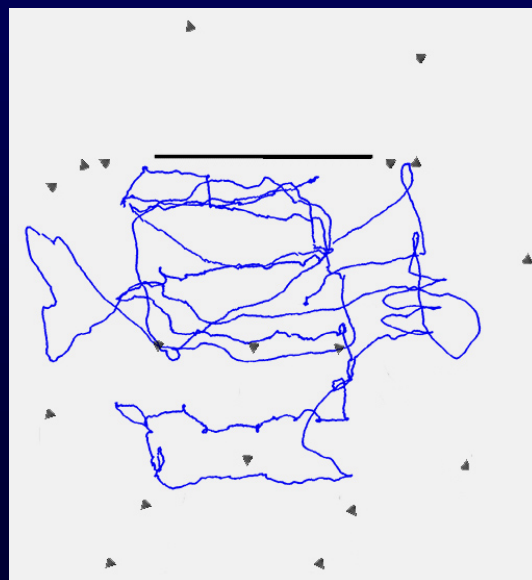
Camera D

Camera E

Camera F

Large scale pose recovery

- **Consider rays through observations**
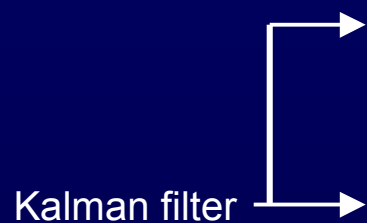- **Rays cross at 3D feature points**

Calibrating wide area cameras

- **Jointly calibrated multiple cameras**
- **Iteratively estimate**
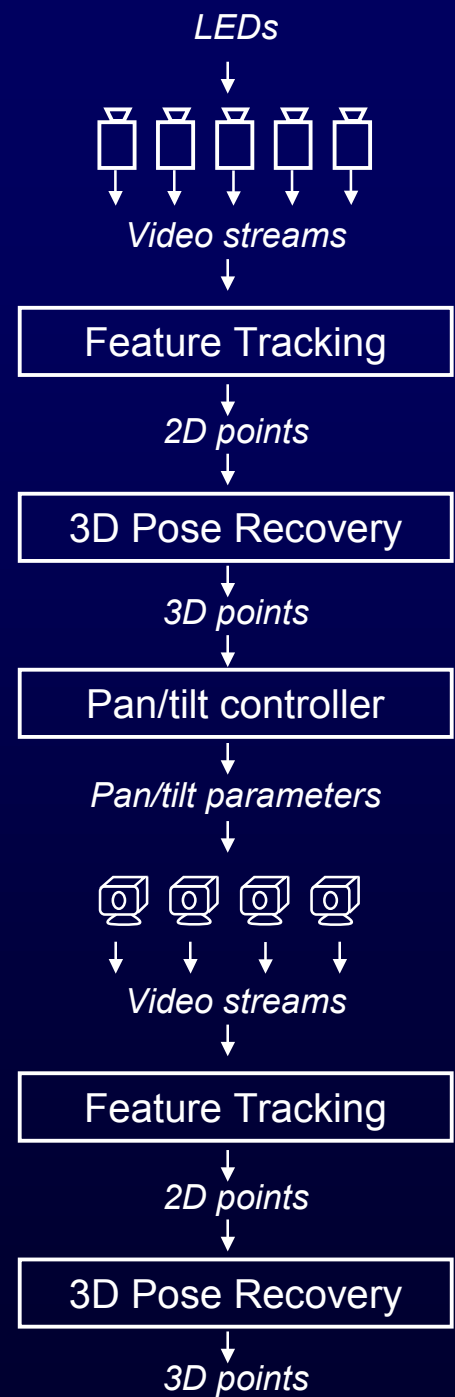 - Camera calibration
  - Target path

Mean Square Projection Error

[Chen, Davis 00]

Unknown correspondence

LEDs

↓

▯ ▯ ▯ ▯ ▯

↓ ↓ ↓ ↓ ↓

Video streams

↓

| Feature Tracking |

2D points

↓

| 3D Pose Recovery |

3D points

↓

| Pan/tilt controller |

Pan/tilt parameters

↓

↓ ↓ ↓ ↓

Video streams

↓

| Feature Tracking |

2D points

↓

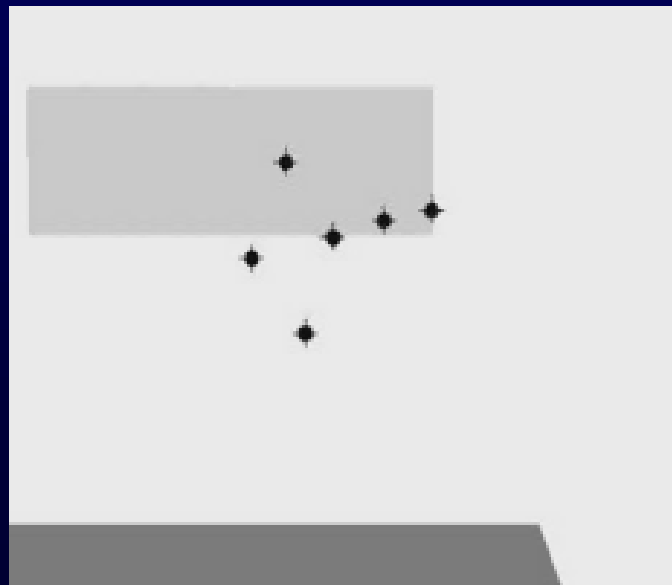| 3D Pose Recovery |

3D points

Kalman filter

Unknown temporal correspondence →

Unknown temporal correspondence

- **Multiple 3D features recovered**
- **Which feature is the head?**
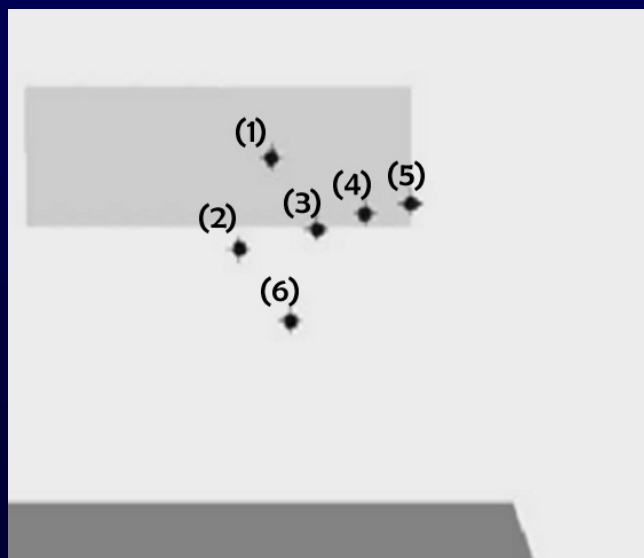- **Each frame is independently derived**

Dynamic motion model

- ## Not single frame triangulation

- ## Dynamic motion model
  - Model continuous motion
  - Update on each observation
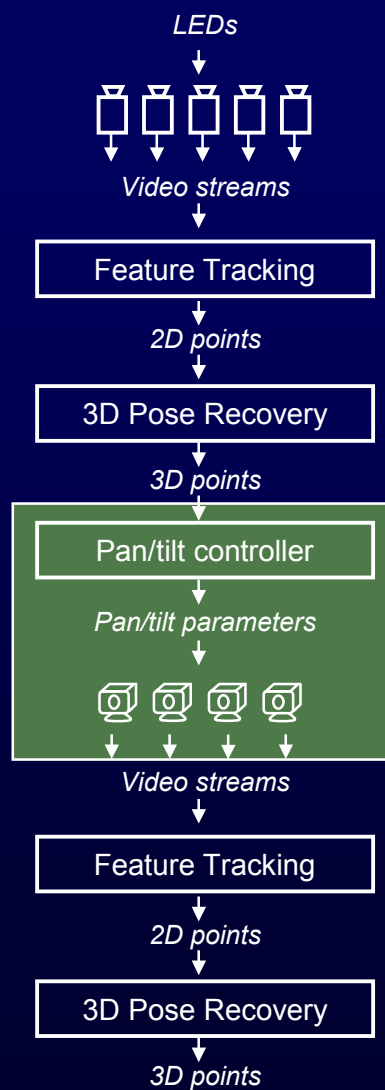 - Estimate position/velocity
  - Extended Kalman filter

[Kalman 60] [Broida86] [Welch97]
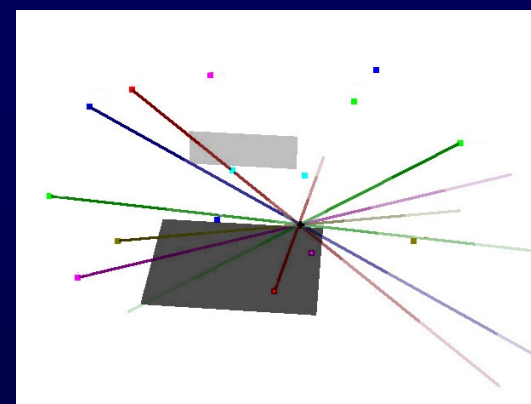
Benefits of motion model

- **Feature IDs maintained**
- **Robust to short occlusion**
- **Synchronized cameras unnecessary**

Sub-region selection

LEDs

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Pan/tilt controller

Pan/tilt parameters

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Simplistic camera model

LEDs

↓

Video streams

↓

Feature Tracking

↓

2D points

↓

3D Pose Recovery

↓

3D points

↓

Pan/tilt controller

↓

Pan/tilt parameters

↓

Video streams

↓

Feature Tracking

↓

2D points

↓

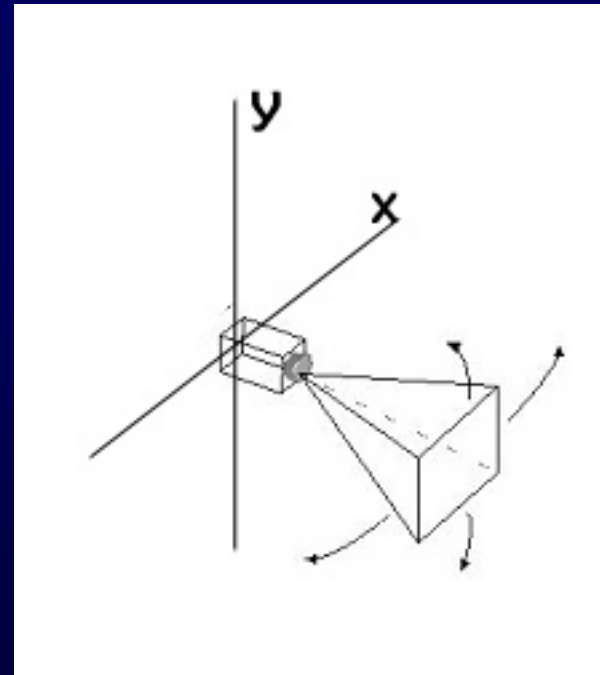3D Pose Recovery

↓

3D points

P/T camera model →   Incorrect camera model →
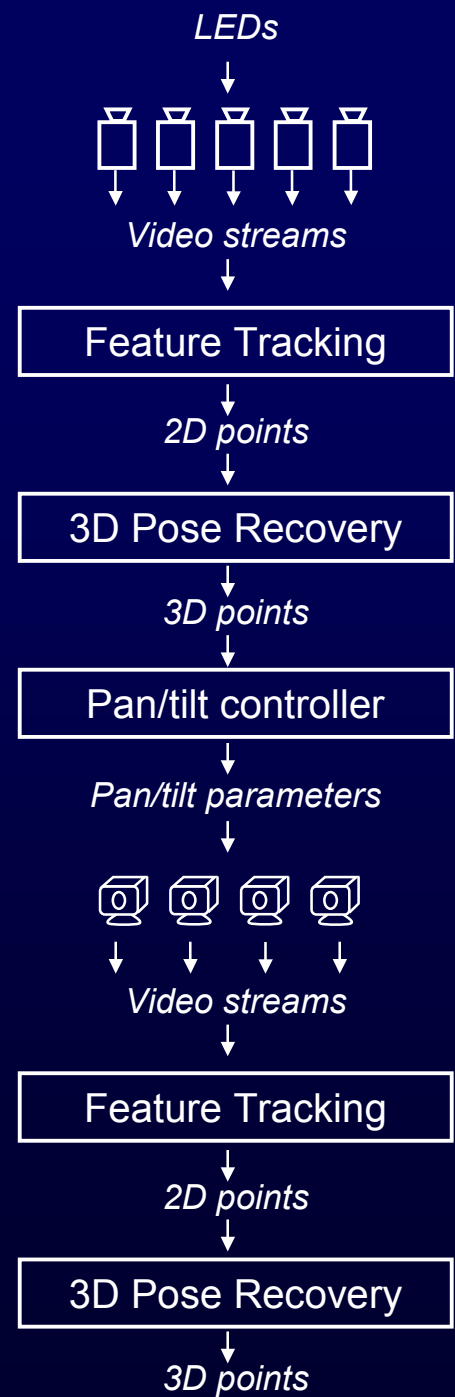
Simplistic camera model

- **Pan/tilt axes not aligned with optical center**

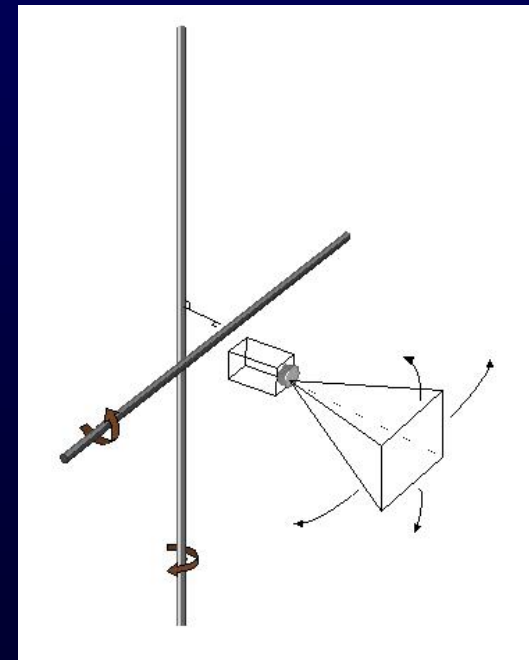$$\begin{bmatrix} I_x \\ I_y \end{bmatrix} = C\ R_y\ R_x \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

New camera model

- **Arbitrary pan/tilt axes**
- **Jointly calibrate axes and camera**
  - Observe known points from several pan/tilt settings
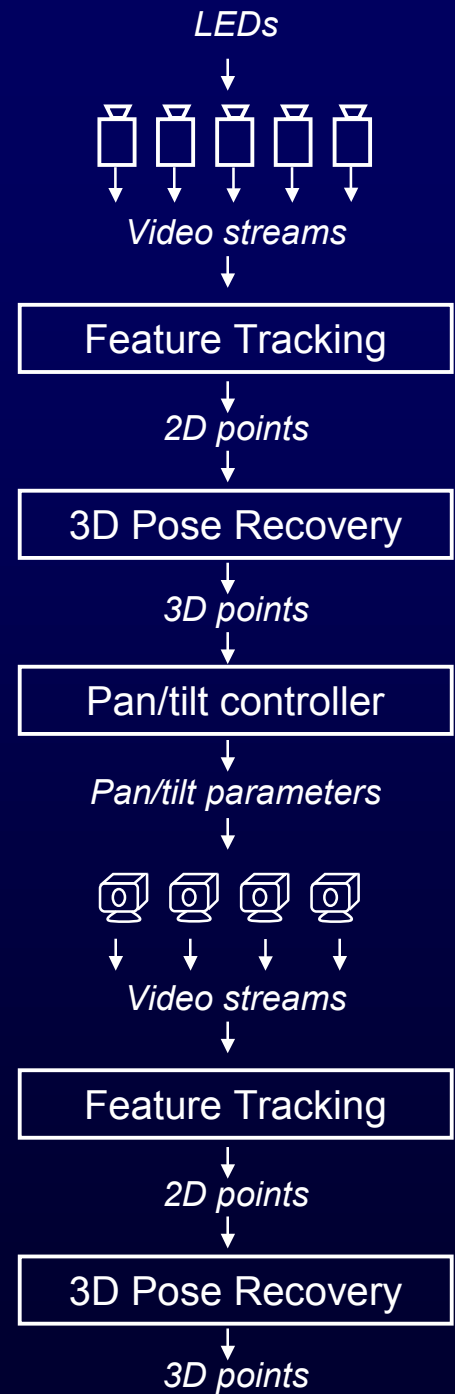 - Fit data with minimum error

$$\begin{bmatrix} \mathbf{I_x} \\ \mathbf{I_y} \end{bmatrix} = \mathbf{C} \; \mathbf{T}_{pan} \; \mathbf{R}_{pan} \; \mathbf{T}_{pan}^{-1} \; \mathbf{T}_{tilt} \; \mathbf{R}_{tilt} \; \mathbf{T}_{tilt}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

[Shih 97]

Latency

LEDs

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Pan/tilt controller

Pan/tilt parameters

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Prediction →

Latency →

Camera motor latency

- **Empirically found 300ms latency**
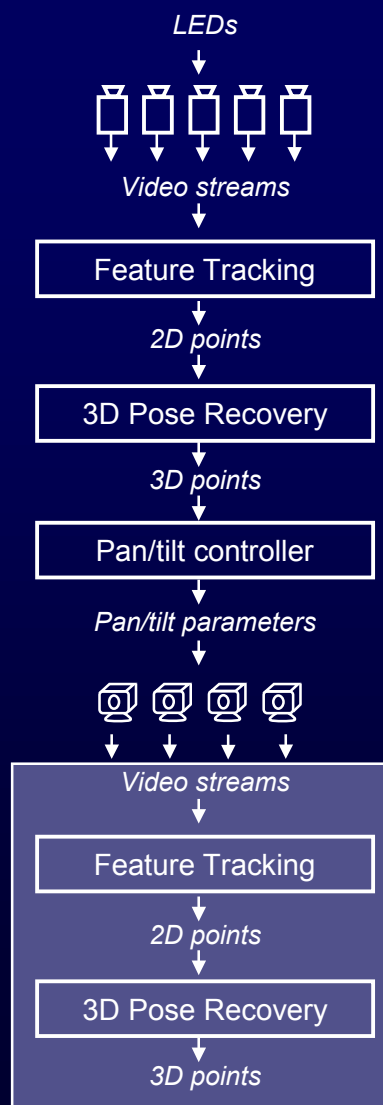- **High velocity targets leave frame**
- **Prevents accurate sub-region selection**

Target motion prediction

- **Predict future target motion**
- **Point camera at predicted target location**
- **Use previous motion model**
- **High velocity objects successfully tracked**
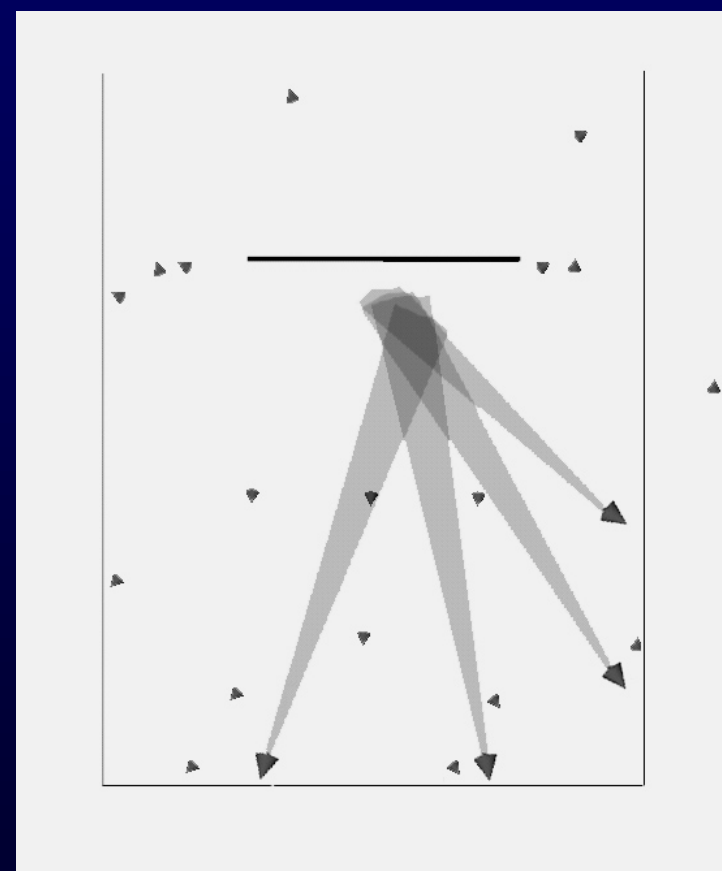
$$P' = P_i + \Delta t \cdot V_i$$

Small scale system

LEDs

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Pan/tilt controller

Pan/tilt parameters

Video streams

Feature Tracking

2D points

3D Pose Recovery

3D points

Small scale physical arrangement

- **4 Pan-tilt cameras point at sub-region**
- **4 SGI O2s digitize video**

Small scale features

- **Painted face features**
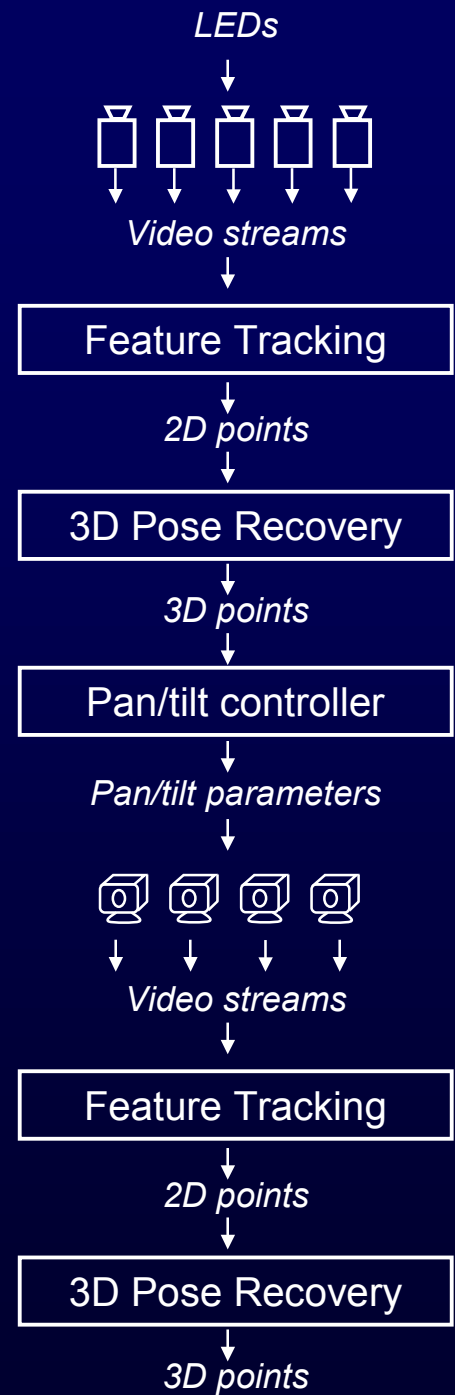- **Image gradient feature tracking**
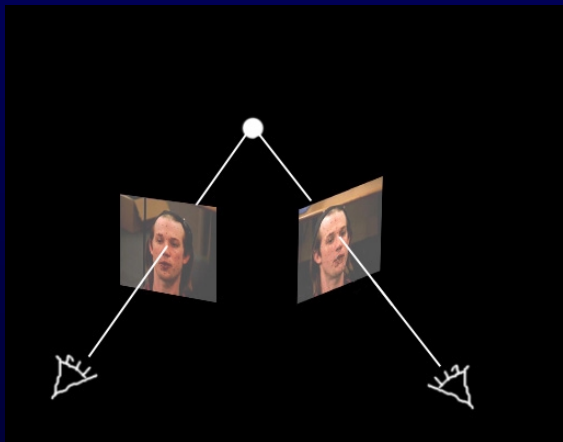
[Lucas,Kanade 81] [Tomasi, Kanade 91]

Small scale pose recovery
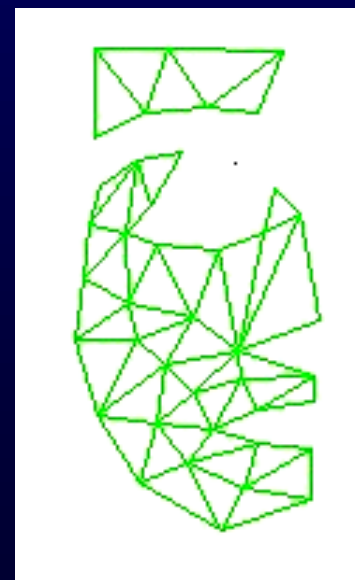
Problems with face recovery

Problems with face recovery

- **Self occlusion**
  - Many points not visible

- **Camera motion not known precisely**
  - Difficult to merge more than two views
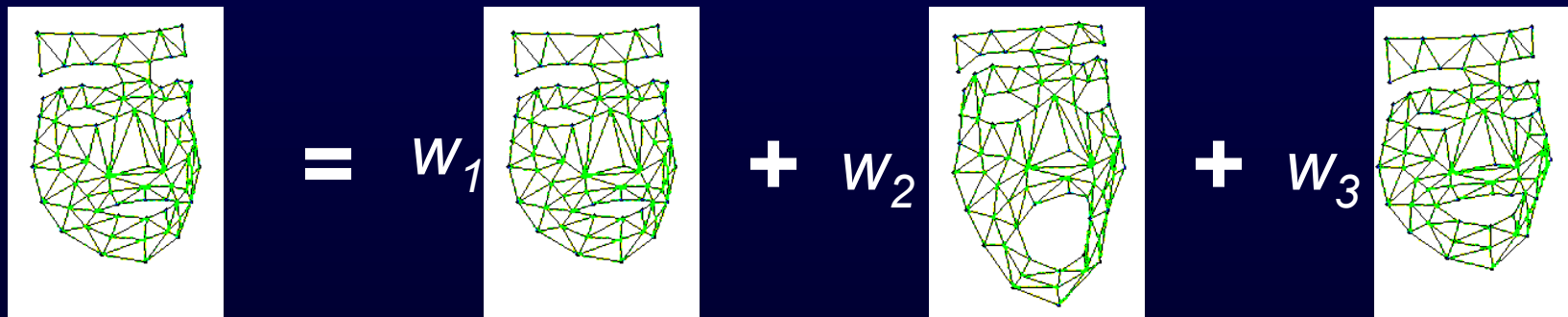

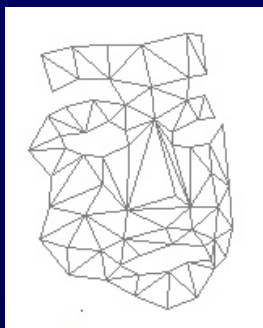View from one camera

Recovered 3D geometry

43

Face model

- **Face model defines the set of valid faces**
  - Linear combination of basis faces
  - Capture basis set under ideal conditions
  - Basis transformation

$$F = \sum w_i B_i$$

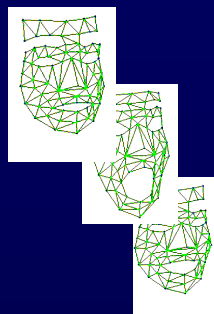$= w_1$ $+ w_2$ $+ w_3$

[Turk, Pentland 91] [Blanz,Vetter 99] [Guenter et.al. 98]

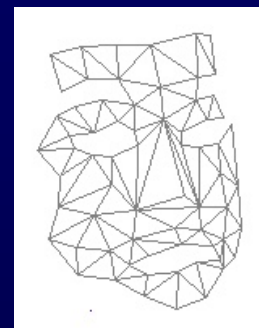# Model evaluation

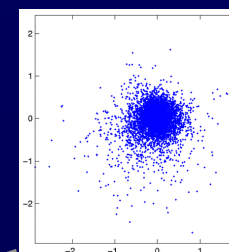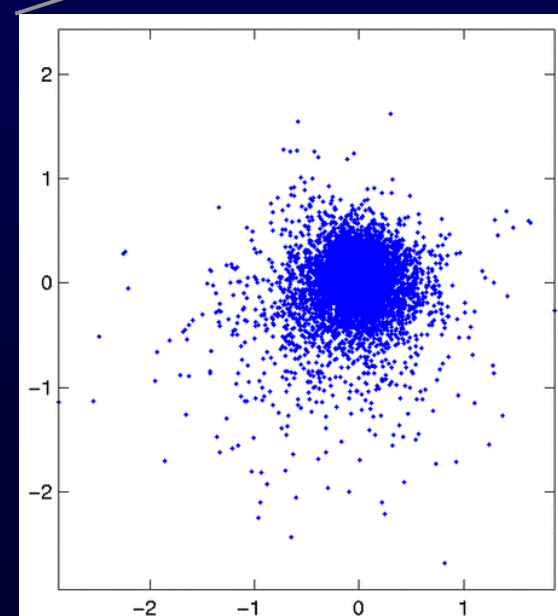Observe everything → Remove features → Fit to model → Reconstruct features → Evaluate error

Mean error < 1.5 mm

Reconstructed face

- **Fit partial data to the model**
- **Use model to reconstruct complete geometry**

Reconstructed geometry from model
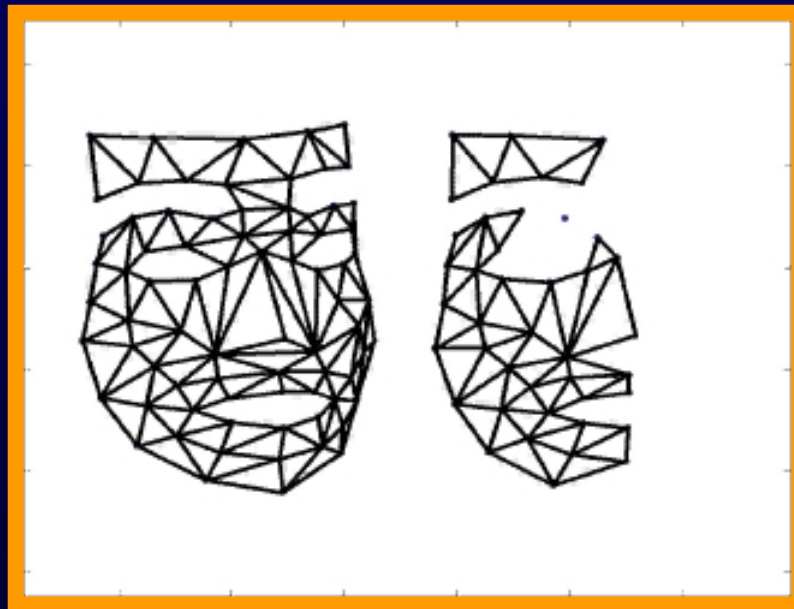
Recovered geometry from video

End to end video

Talk outline

- **Introduction**
- **Framework**
  - Hierarchical paradigm
  - Data driven analysis
  - Model based solutions
- **System implementation**
  - Large scale recovery
  - Sub-region selection
  - Small scale recovery
- **End-to-end video**
- **Summary and future work**

Summary of contributions

- **Framework for mixed scale motion recovery**

  - Hierarchical paradigm

  - Data driven analysis

  - Model based solutions

- **Specific system design**

  - High resolution/range ratio

  - Scalable

  - Robust to occlusion

  - Automated

- **Application to simultaneous face-body capture**

Future directions

- **Application to other domains**
- **More levels of hierarchy**
- **Selection of multiple sub-regions**
- **Alternate system designs**

Acknowledgements

Prof. Pat Hanrahan, Prof. Brian Wandell, Prof. Chris Bregler, Prof. Gene Alexander, Prof. Marc Levoy, Cindy Chen, Ada Glucksman, Heather Gentner, Homan Igehy, Venkat Krishnamurthy, Tamara Munzner, François Guimbretière, Szymon Rusinkiewicz, Maneesh Agrawala, Lucas Pereira, Kari Pulli, Shorty, Sean Anderson, Reid Gershbein, Philipp Slusallek, Milton Chen, Mathew Eldridge, Natasha Gelfand, Olaf Hall-Holt, Humper, Brad Johanson, Sergey Brin, Dave Koller, John Owens, Kekoa Proudfoot, Kathy Pullen, Bill Mark, Dan Russel, Larry Page, Li-Yi Wei, Gordon Stoll, Julien Basch, Andrew Beers, Hector Garcia-Molina, Brian Freyburger, Mark Horowitz, Erika Chuang, Chase Garfinkle, John Gerth, Xie Feng, Craig Kolb, Toli, Mom, Dad, Holly Jones, Chris, Crystal, Lara, Grace Gamoso, Matt Hamre, Nancy Schaal, Aaron Jones, Bandit, Xiaoyuan Tu, Abigail, Shefali, Liza, Phil, Deborah, Brianna, Alejandra, Miss Dungan, Gabe, Sedona, Sharon, Gonzalo, and many other children whose names I can no longer remember