# 3D head pose estimation from multiple distant views

X. Zabulis, T. Sarmis, A. A. Argyros

Institute of Computer Science, Foundation for Research and Technology – Hellas (ICS-FORTH)

The problem

- Estimate *X*, *Y*, *Z* (location) and *yaw*, *pitch* and *roll* (pose), from multiple calibrated <u>distal</u> views

- Based on face detection

- Face detection, in distal viewing, is a significant component of the problem
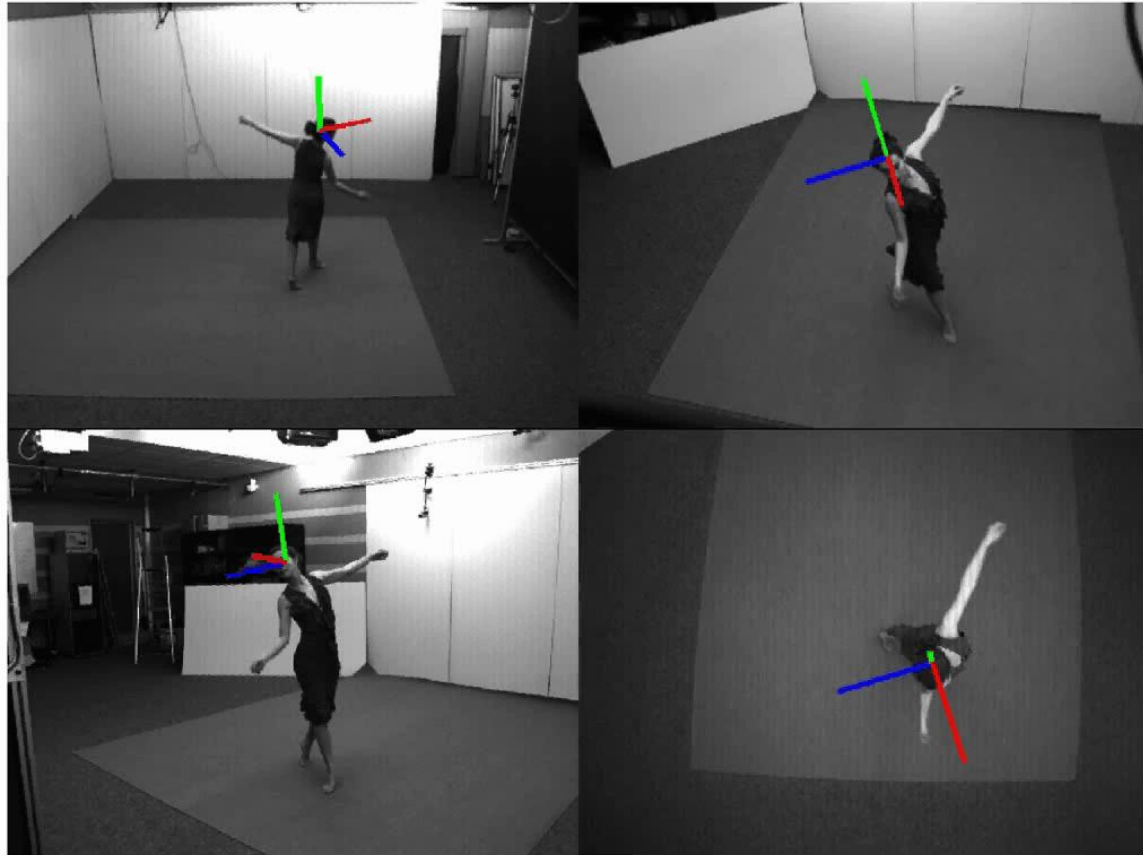
**Reference dataset from INRIA (IXMAS)**

Past solutions and issues

- Templates, detector arrays
- Nonlinear regression, manifold embedding
- Tracking rigid landmarks
- Face alignment
- Fusion of single view estimates
  - Bayesian, Conditional probability, Joint likelihood
  - Only coarse pose estimation is reported

Single view

Proposed method

- Textured visual hull to locate the head in 3D

- Unfold head texture from the visual hull, on a hypothetical sphere around the head

- Face detection in the unfolded spherical image is <u>simple</u> and <u>yields pose estimation</u>

- The spherical image <u>collects all available facial texture fragments</u> to increase the support for face detection, <u>despite occlusions</u>
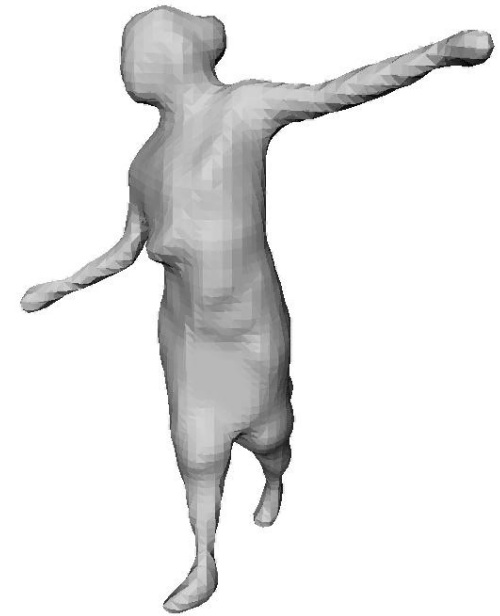
Visual Hull
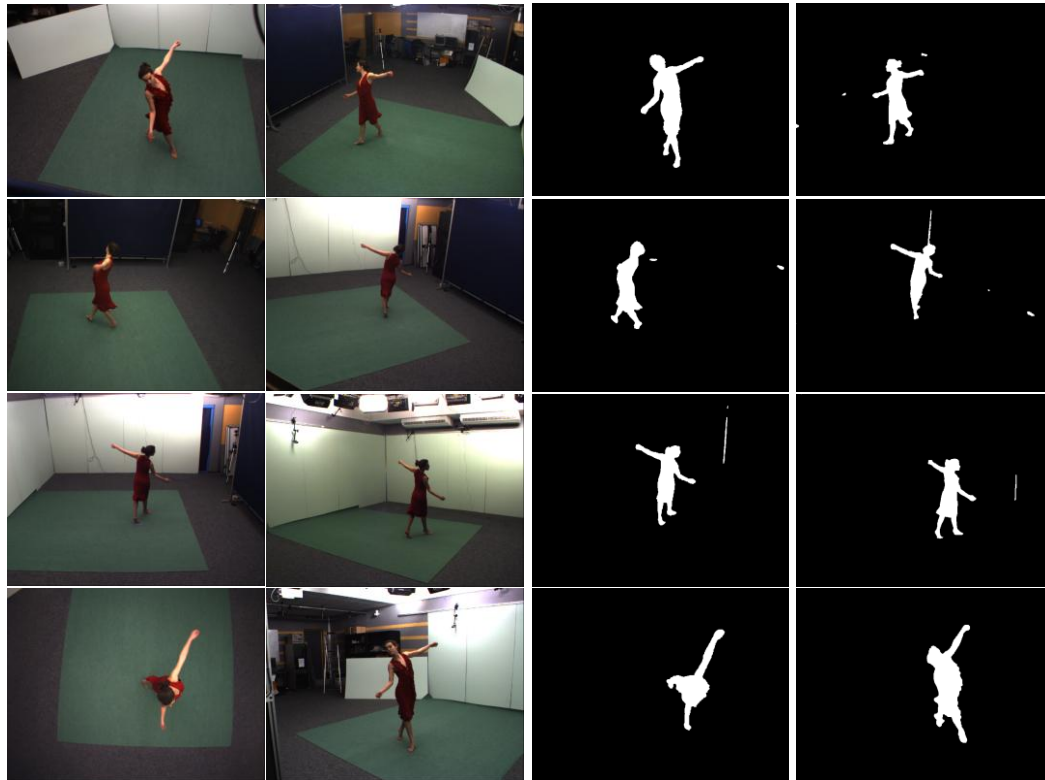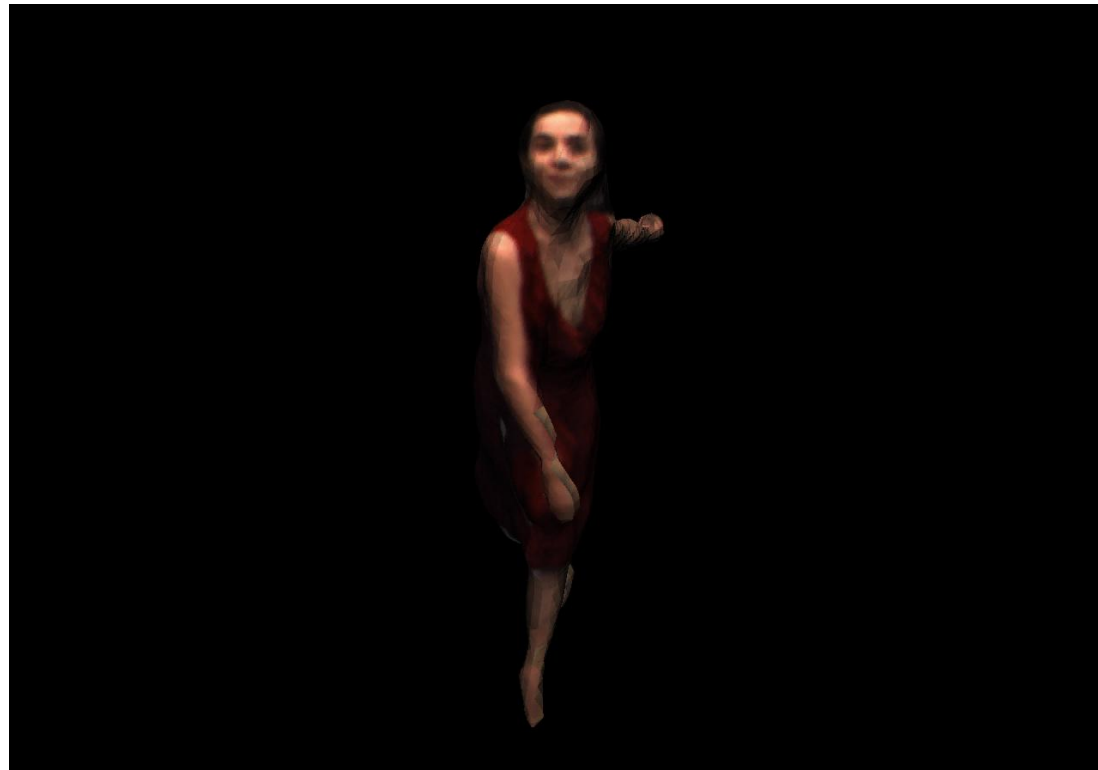

Original images   Background subtracted   Visual Hull

Textured visual hull

- Z-buffer to determine the view(s) that each mesh triangle is visible from

- Blending of textures when multiple views image the same triangle
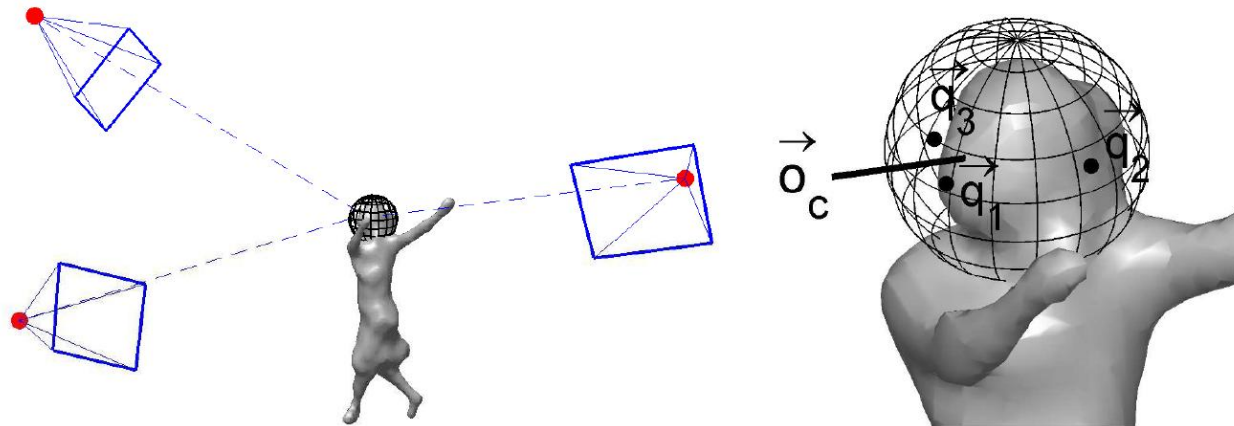
Head localization

- 3D variant of Mean–Shift :
  - spherical kernel registered with head
  - minimize sum of distances of head points from the spherical kernel's surface
- Initialization:
  - location estimate in the previous frame
 - 1st frame: face detection(s) in the original images

Head center
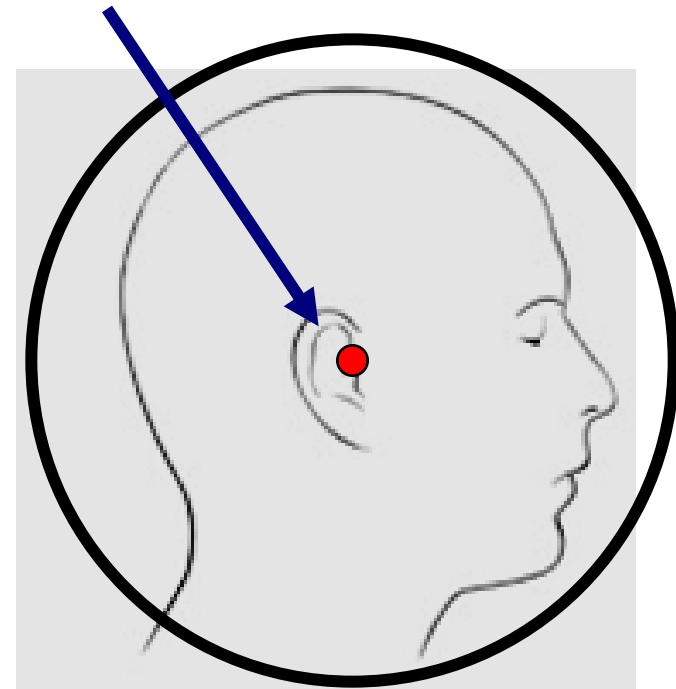
Coarse orientation estimation

- Face detections at head projections define lines intersecting a hypothetical sphere around the head
- Vectors from head center to these intersection are <u>robustly</u> combined into an orientation estimate
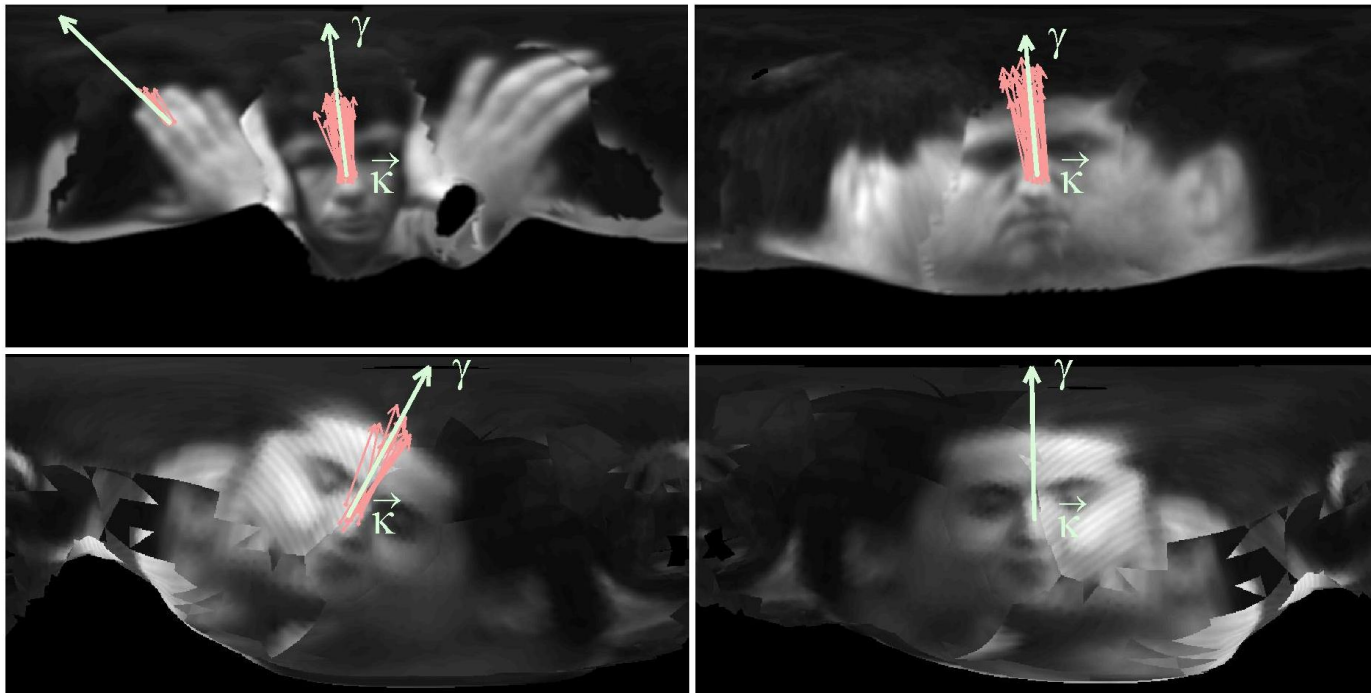
Head pose estimation I

- Let a hypothetical sphere around the head
- Center-project the visual hull texture on this sphere
- *The formed spherical image contains <u>exactly one frontal</u> face appearance and collects texture from all views*
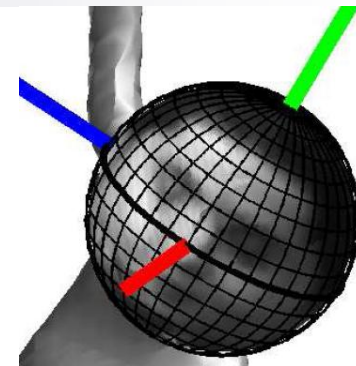- Detecting the face in it provides a head pose estimate
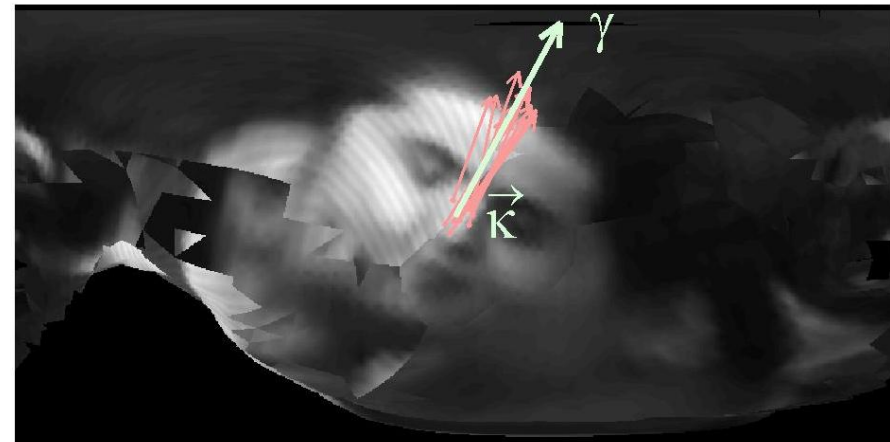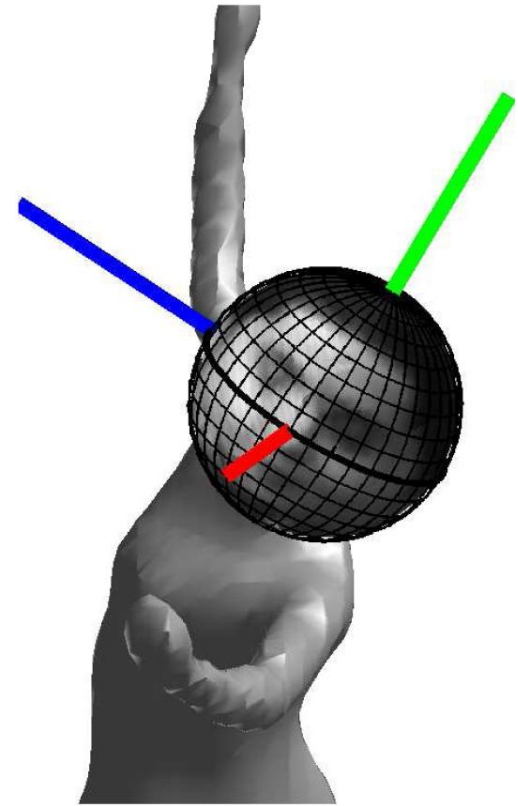
Head pose estimation II

- Generic frontal face detector

- Face center yields pitch and yaw

- Query all 2D poses to find roll component

Head pose estimation III

- Spherical parameterization causes distortions if the face projects near a pole

- Using the coarse orientation estimate, the sphere is center-rotated (in place) and the face projects on the equator

The face detection process is iterated until convergence, availing greater accuracy

Benchmark dataset with ground truth

- A hemisphere of poses
- Roll and pitch rotations
- Mannequin head mounted on tripod with marked rotational gratings

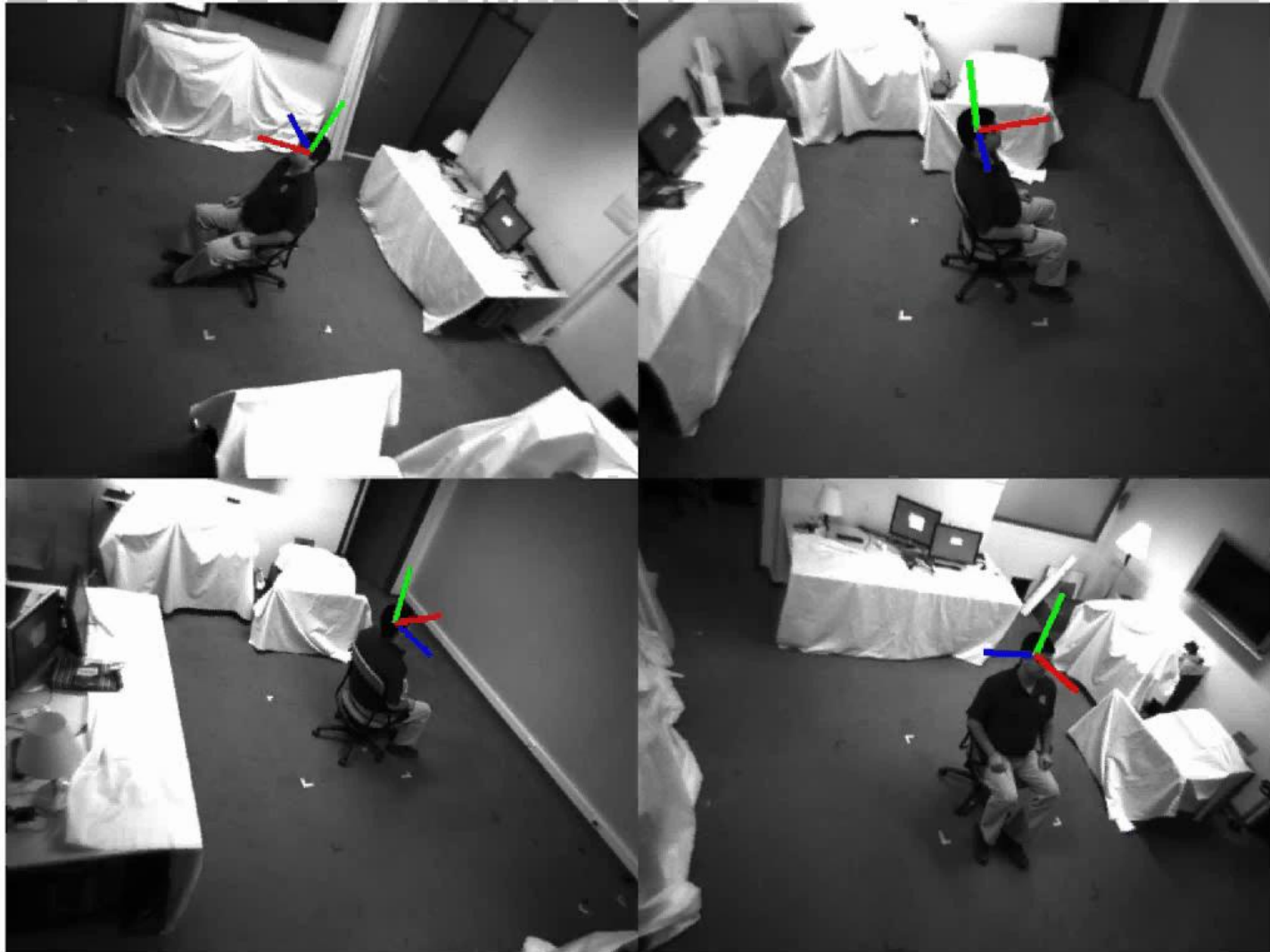- **Available at http://www.ics.forth.gr/cvrl/headpose**
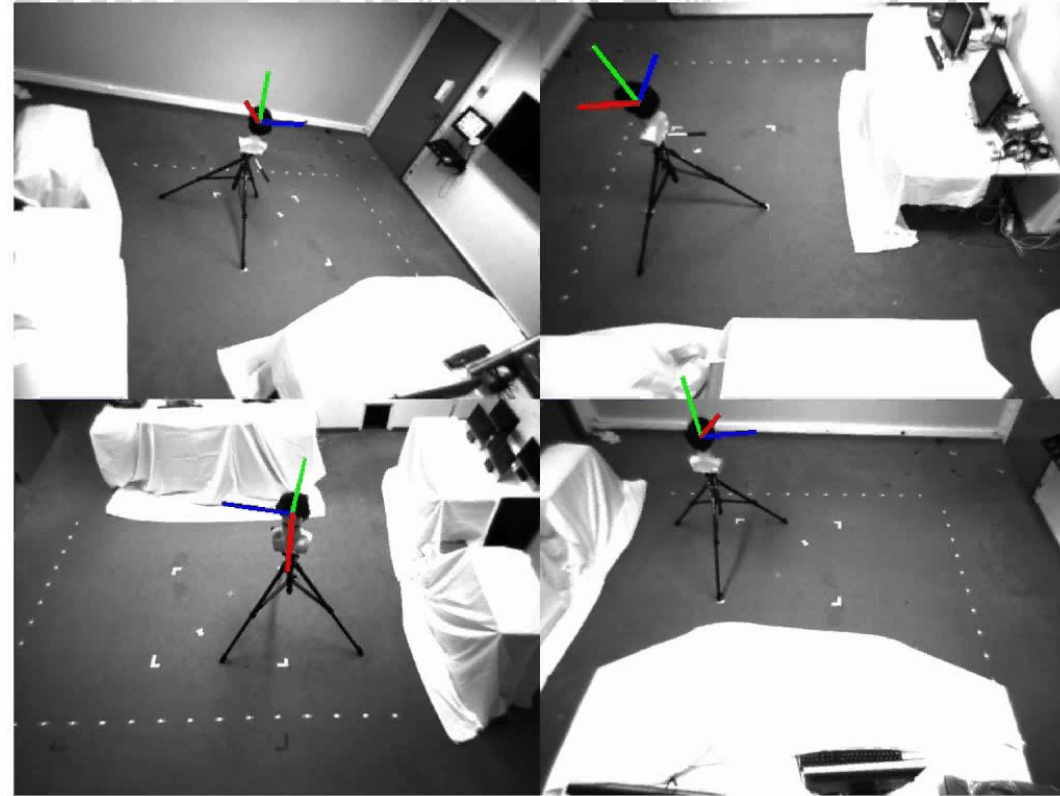
Ground truth experiments

• Improvement over state of the art (~8 - 10degrees), in the distal viewing context

• Average error ~3 degrees

• Graceful degradation: error grows with lack of coverage

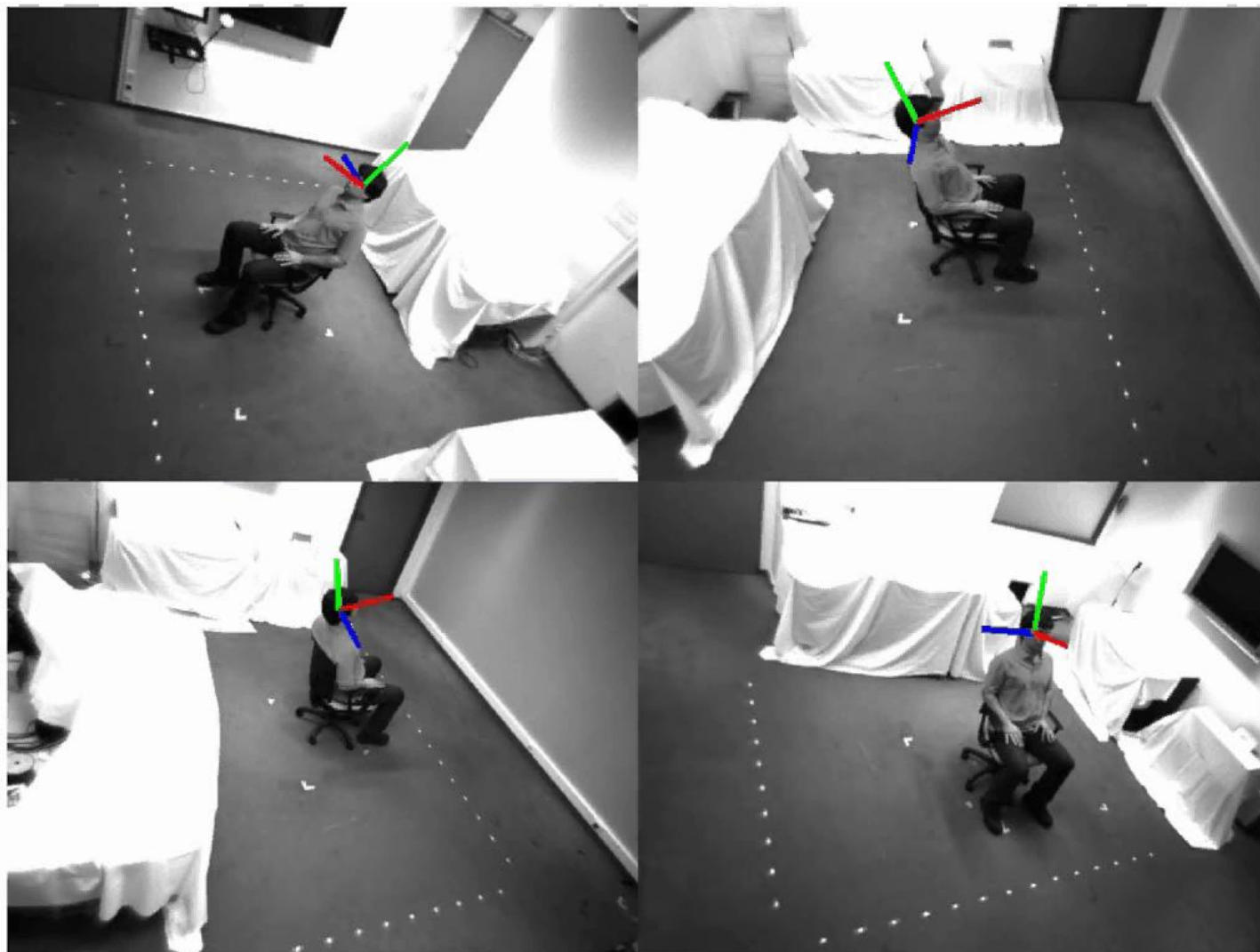| Poses	Yaw	Pitch	Roll
Hemisphere	5.4° (6.1°)	2.8° (2.4°)	8.0° (7.6°)
Usual poses	2.3° (2.1°)	2.1° (1.6°)	5.3° (4.6°)
Pitch	1.9° (1.2°)	2.4° (1.6°)	4.7° (3.9°)
Roll	unknown	3.5° (3.0°)	6.7° (5.3°)

YAW error   PITCH error   ROLL error
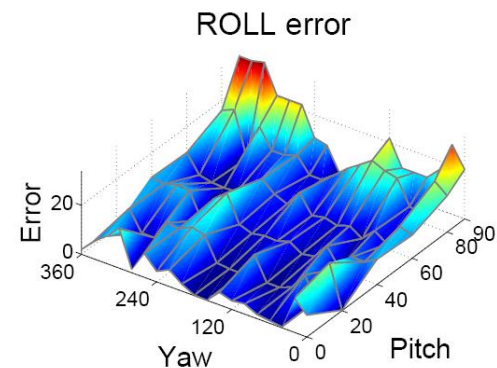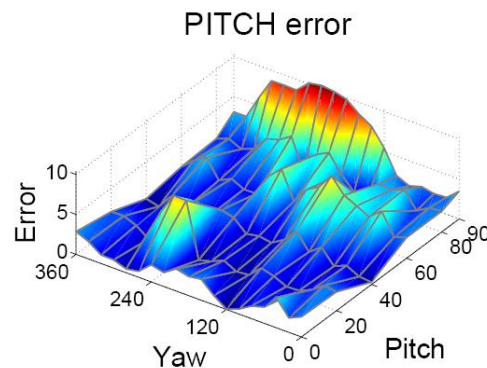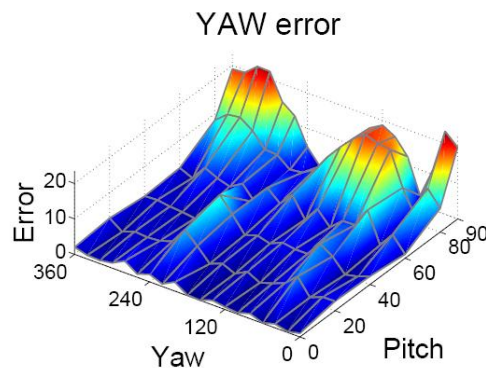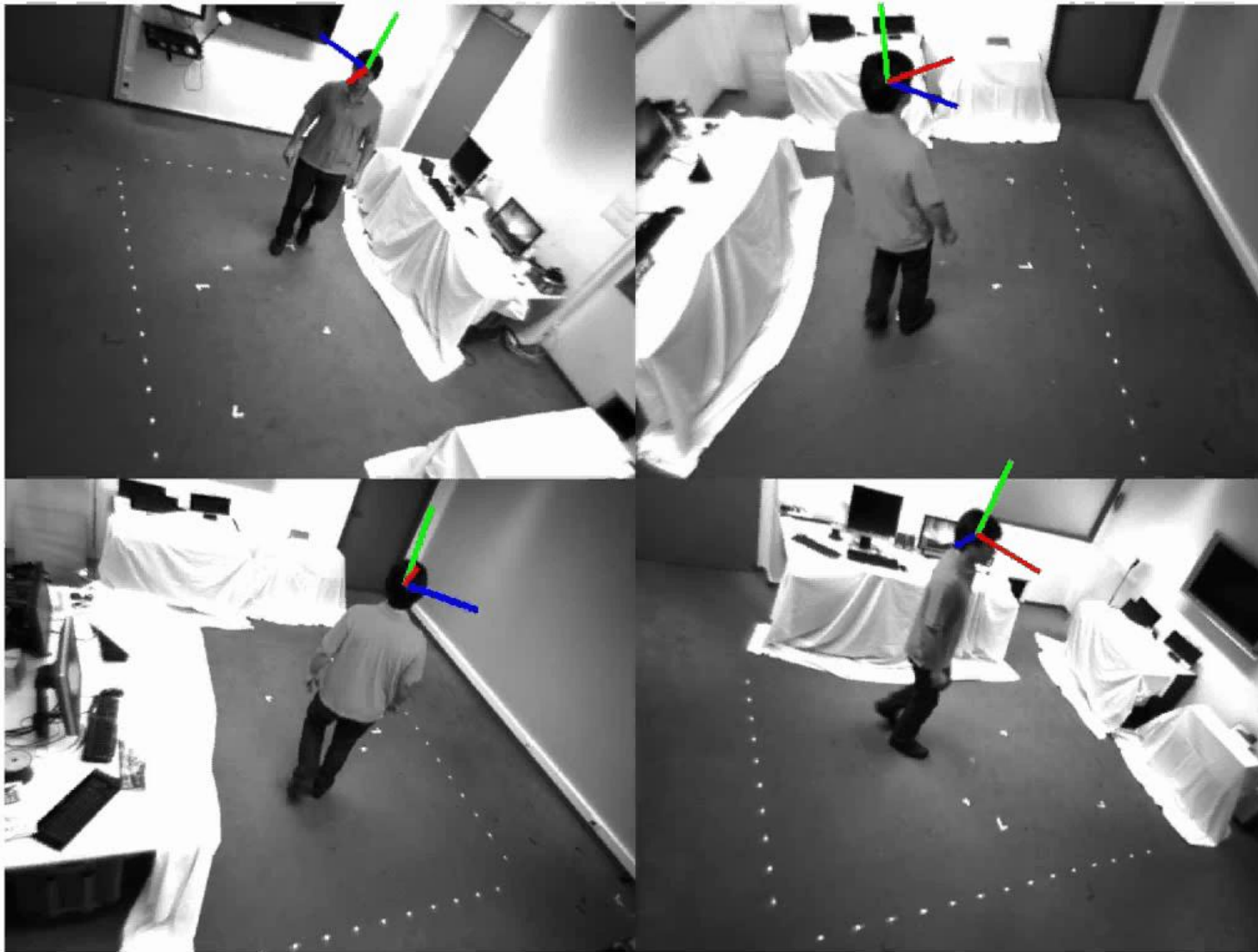
Experiments (chair 360)

No tracking of filtering was applied in these results. Raw estimator output shown.

Experiments (hands)

No tracking of filtering was applied in these results. Raw estimator output shown.

Experiments (standing)

No tracking of filtering was applied in these results. Raw estimator output shown.

Experiments (occlusions)

No tracking of filtering was applied in these results. Raw estimator output shown.

Experiments (2persons)

No tracking of filtering was applied in these results. Raw estimator output shown.

Future work (in progress)

- Apply robust tracking framework
- Optimize
  - GPU implementation
  - Compute only the relevant (face) texture

Q & A

Thank you for your attention