# FI-CAP: ROBUST FRAMEWORK TO BENCHMARK HEAD POSE ESTIMATION IN CHALLENGING ENVIRONMENTS


Sumit Jha, Carlos Busso

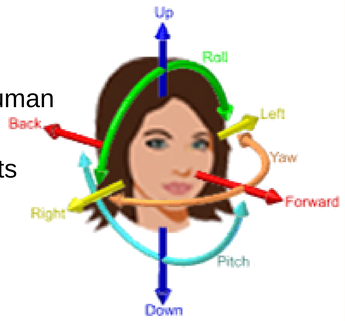
UT Dallas MSP Multimodal Signal Processing Laboratory

THE UNIVERSITY OF TEXAS AT DALLAS

**Erik Jonsson School of Engineering & Computer Science** at the University of Texas at Dallas, Richardson, Texas 75080, USA

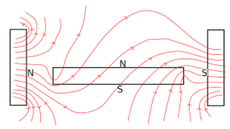

## MOTIVATION

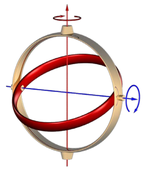
- Head Pose Estimation in the wild is an important problem to solve
- Naturalistic Driving behavior
  - Visual attention
  - Infotainment
- Human-Human and Human Computer Interaction
  - Interaction with robots
  - Emotion
- Smart Interfaces
  - Virtual reality
  - Wake up cues
- Deep learning can provide reliable methods to estimate head pose estimation across environment
  - Require large databases with reliable annotation
- Solution : Fiducial markers based design that can provide reliable continuous head pose annotation



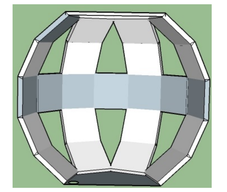

## AVAILABLE OPTIONS

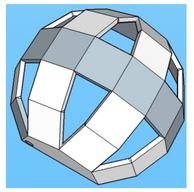
- Directional Suggestion [Gourier et al. 2004, Savran et al. 2008]
  - Ask the subjects to look at particular directions
    - Or point with a laser attached to the head [Rae and Ritter 1998]
  - Limitations: only suggestive not quantitative
- Magnetic Sensors [Ba and Odobez 2005, Ariz et al. 2016]
  - Position and orientation based on magnetic field produced by a transmitter
  - Sensors placed on head that calculates the magnetic forces acting on it
  - Limitations: environment cannot have magnetic material
- Inertial Measurement Units
  - IMUs calculate the direction and magnitude of various forces being acted on a body [Morency et al. 2003, Tawari et al. 2014]
  - Placed on head to capture forces during head movement
  - Limitations: Susceptible to small movements and vibrations
- Motion Capture [Busso et al. 2008, Schwarz et al. 2017]
  - Highly reflective surfaces or LEDs placed on head and faces
  - Limitations: Highly controlled and expensive setup required

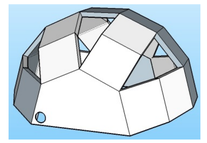


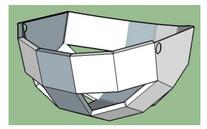



## Fi-Cap DESIGN









- 3D design with horizontal and vertical fiducial markers
  - Multiple tags visible at all head poses
- Large tag size (4 cm)
- Requires an additional camera to track the tags

- Robust detection in variety of head poses
- No occlusion of facial features
- Continuous tracking at high resolution

## EVALUATION

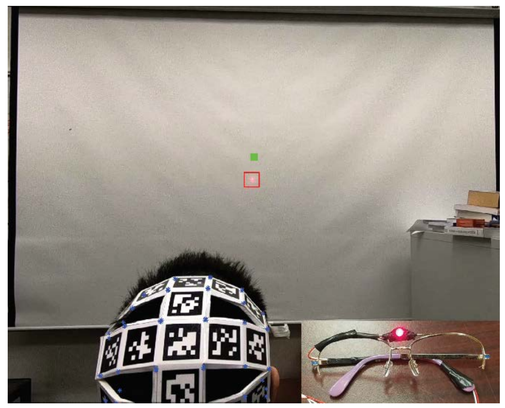
**Rendering in 3D virtual environment**

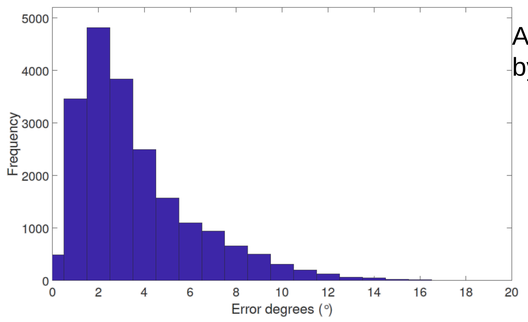
- Virtual design using blender
- Animated with known head movements
- Comparison of given head movements and estimations using the cap

| Data | Mean Error [°] | Median Error [°] | 95 percentile Error [°] |
|---|---|---|---|
| 540p neutral | 1.17 | 1.18 | 1.95 |
| 540p high | 1.45 | 1.61 | 2.10 |
| 540p low | 1.17 | 0.95 | **2.87** |
| 1080p neutral | 0.39 | 0.36 | 0.66 |
| 1080p high | 1.12 | 1.15 | 1.46 |
| 1080p low | 0.93 | 1.01 | 1.30 |

**Controlled settings in the lab**

- To validate the performance on real-world data
- Participants asked to wear glasses with Laser attached at the center
- Look at a screen such that a laserbeam and the tags can be simultaneously tracked
- Predict the beam location using the Fi-Cap design using limited calibration set



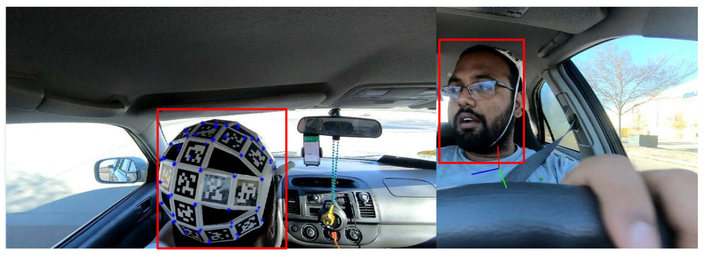



Angular error between beam location obtained by template matching and predicted beam

| Data Calibration | Mean Error [°] | Median Error [°] | 95 percentile Error [°] |
|---|---|---|---|
| First 10 | 1.17 | 1.18 | 1.95 |
| First 100 | 1.45 | 1.61 | 2.10 |
| First 500 | 1.17 | 0.95 | **2.87** |
| Random 100 | 0.39 | 0.36 | 0.66 |

## CONCLUSIONS

- A promising tool to annotate head pose in diverse environments
- Open new research opotunities
  - Evaluate other sensors such as Radar, deep camera
- Our goal is to collect naturalistic driving data to study driver attention
  - Pilot data shows promising result with multiple tags visible in most frames (99.16%)




This work was supported by Semiconductor Research Corporation (SRC) / Texas Analog Center of Excellence (TxACE), under task 2810.014