Learning, Detecting and Localizing 3D Object Classes from Arbitrary Viewpoints

Matthew Toews and Tal Arbel
Centre for Intelligent Machines
McGill University, Montreal, Canada

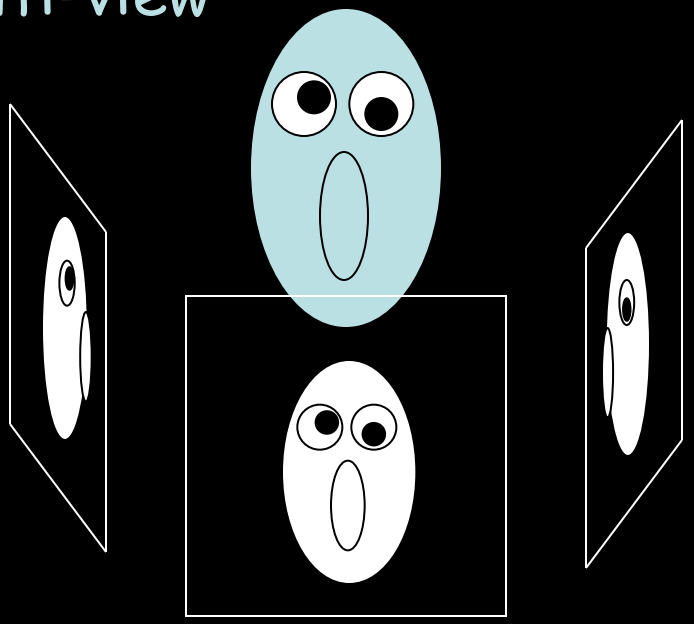
McGill
Centre for
Intelligent Machines

3dRR
2007

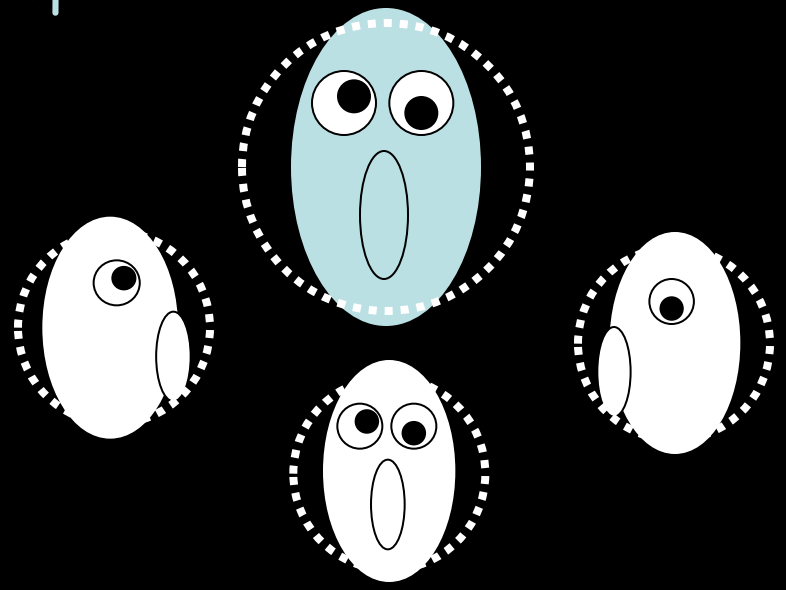
Multi-view vs. Viewpoint-Invariant

- Which to use?

Multi-view



Viewpoint Invariant



Mental Rotation

- Time to decide if two 3D objects are similar is proportional to the angular difference between them

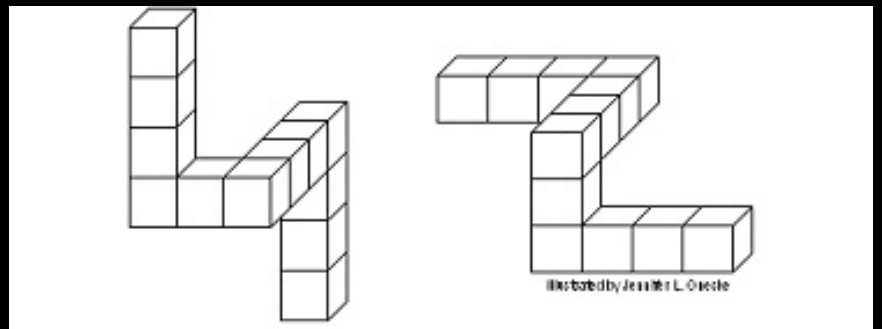


Figure 1: Based on Shepard & Metzlar's 'Mental Rotation Task'

Figure 2: Mental Rotation Task Based on Canonical Orientations

Roger Shepard &
Jacqueline Metzler, 1971

Recognition & Invariance

- Recognizing specific objects seems to be viewpoint-dependent
- Recognizing object classes seems to be viewpoint-invariant

Shimon Edelman
1995

Recognition & Invariance

- Recognizing specific objects seems to be viewpoint-dependent
- Recognizing object classes seems to be viewpoint-invariant

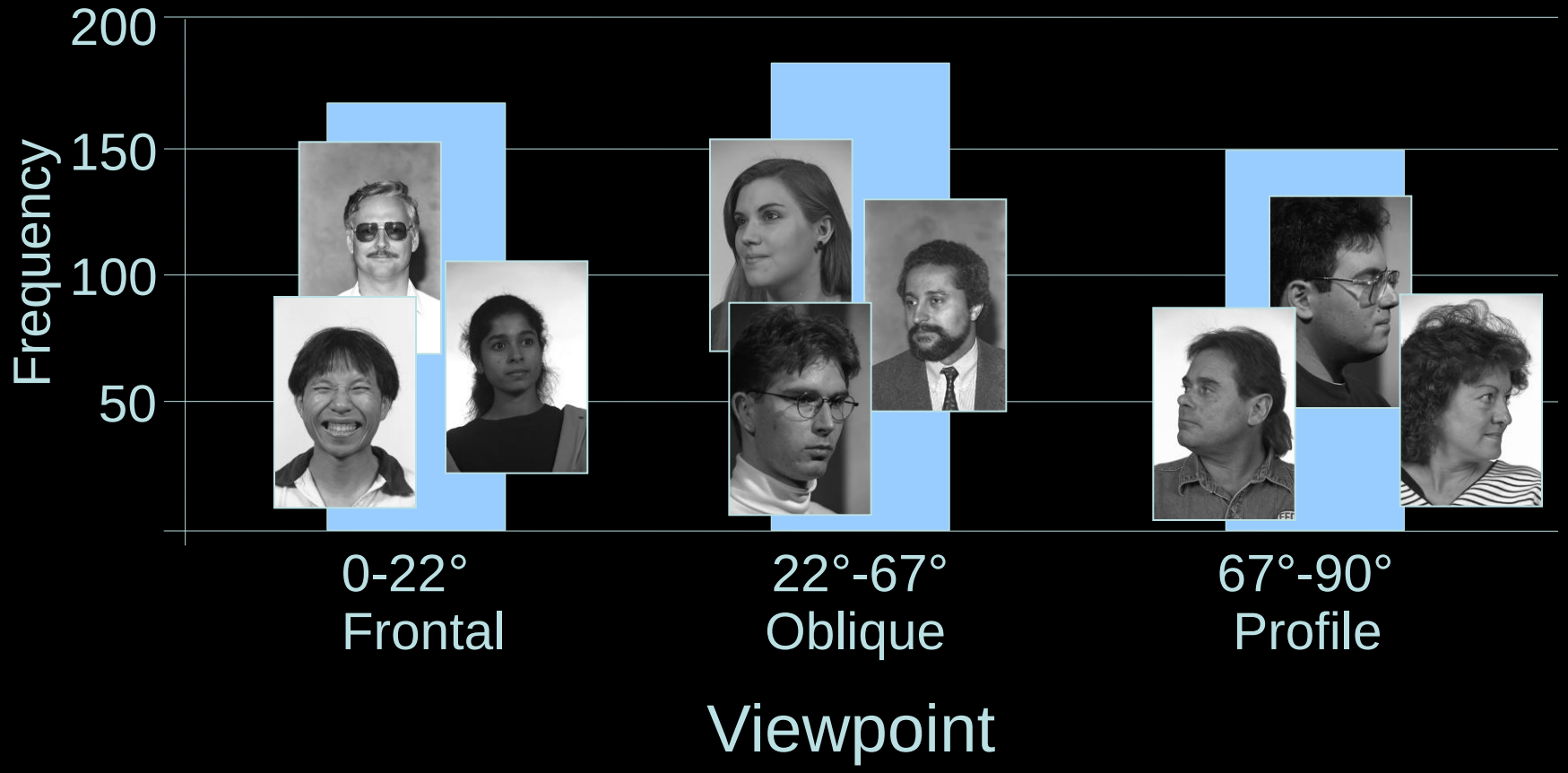
Shimon Edelman
1995

Methodology

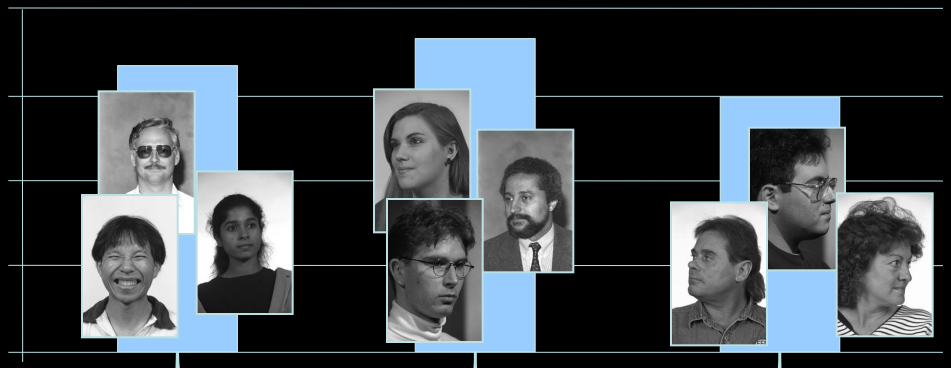
- Learn multi-view and viewpoint-invariant models from the same data.
- Compare detection performance.
- Data: faces, viewpoint variation.

Learning

- Color FERET database
  - 500 unique faces, viewpoint labels

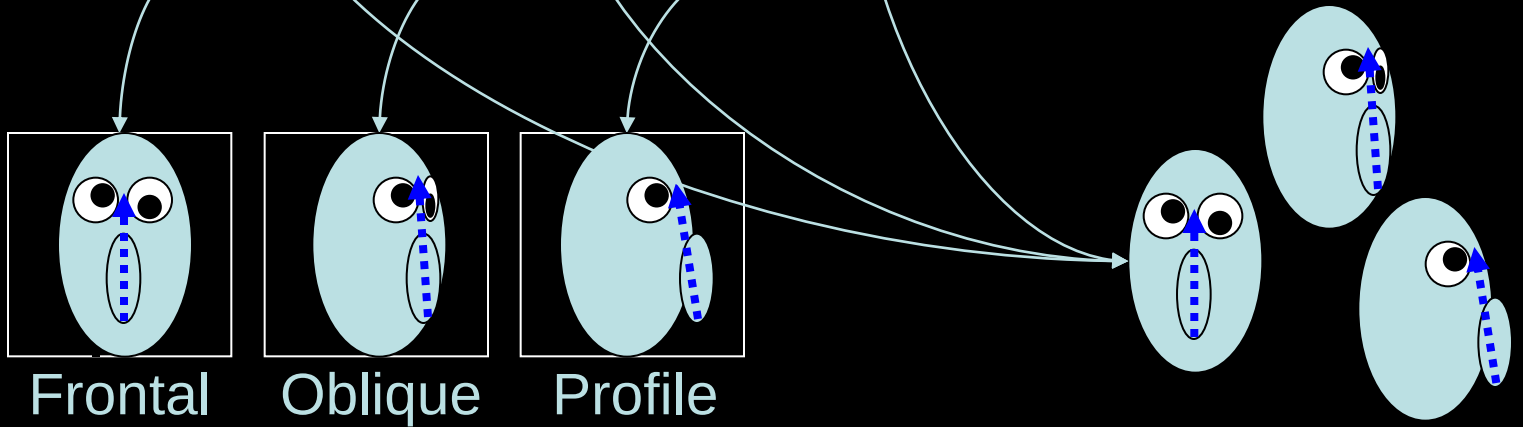


Learning



* Manually labelled OCIs

* Robust feature clustering

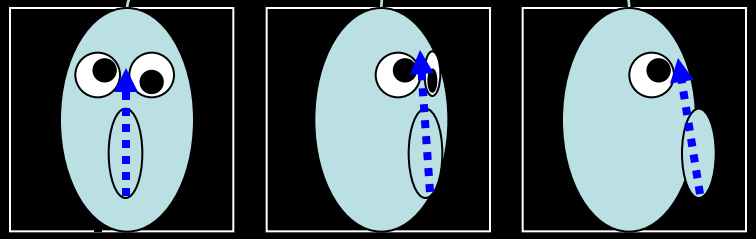


Multi-view Model

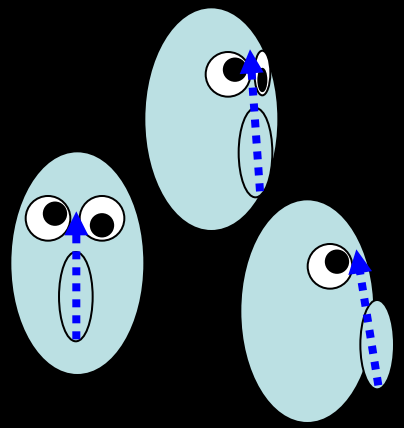
Viewpoint-invariant Model

Detection and Localization

CMU Profile Database
(subset, 97 faces)

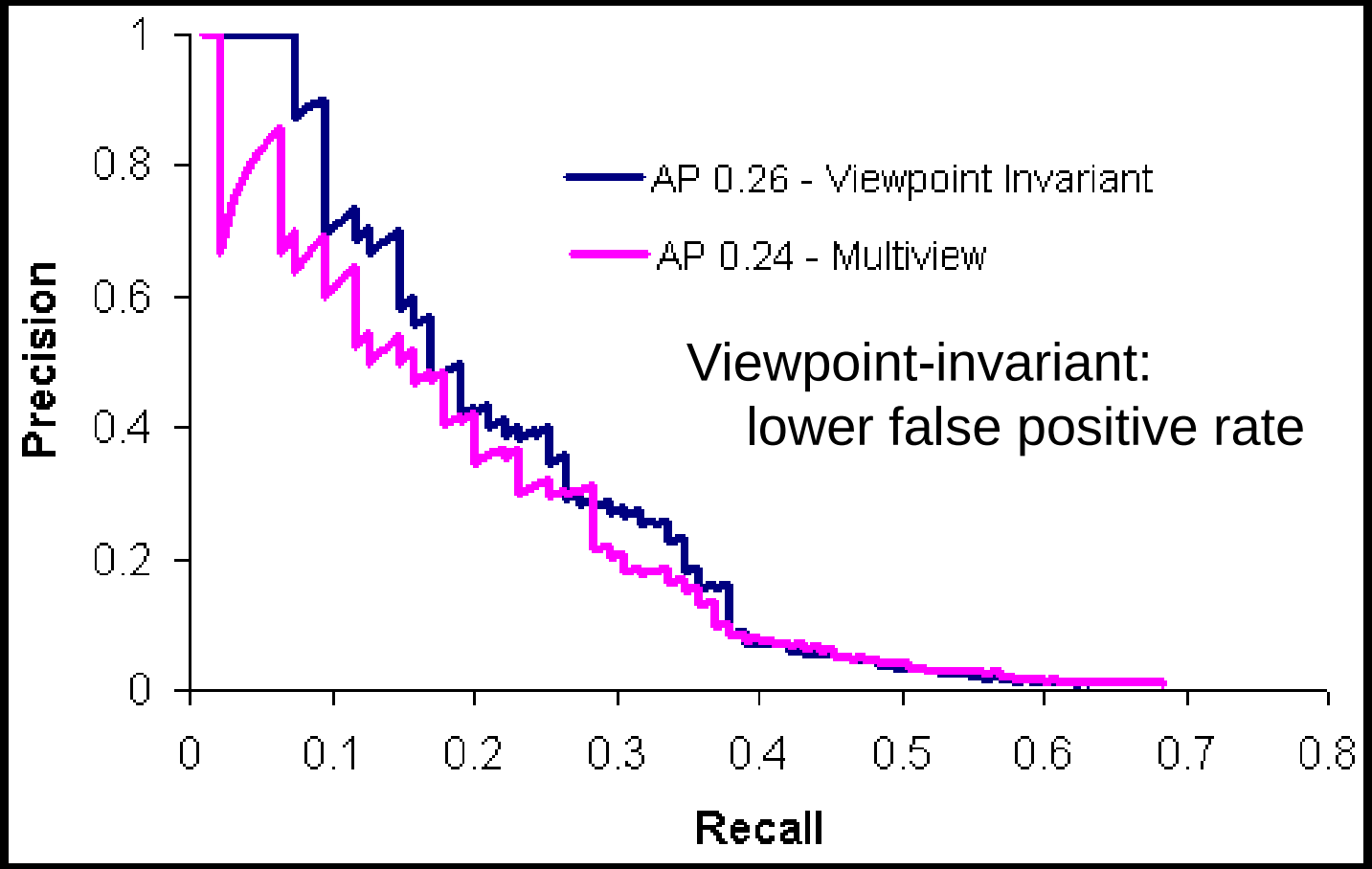


Multi-view

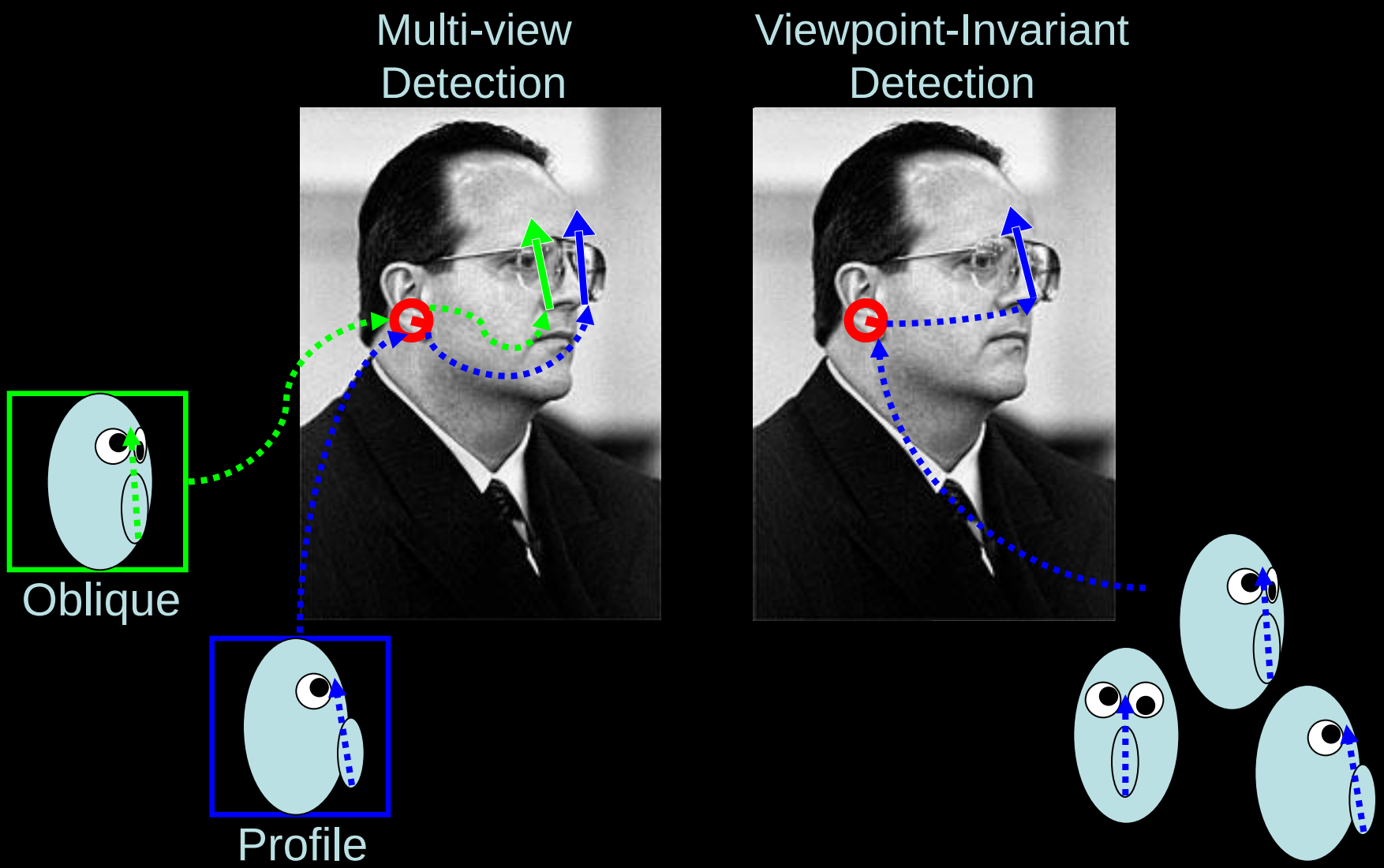


Viewpoint-invariant

Detection Comparison



Example: Localization Ambiguity



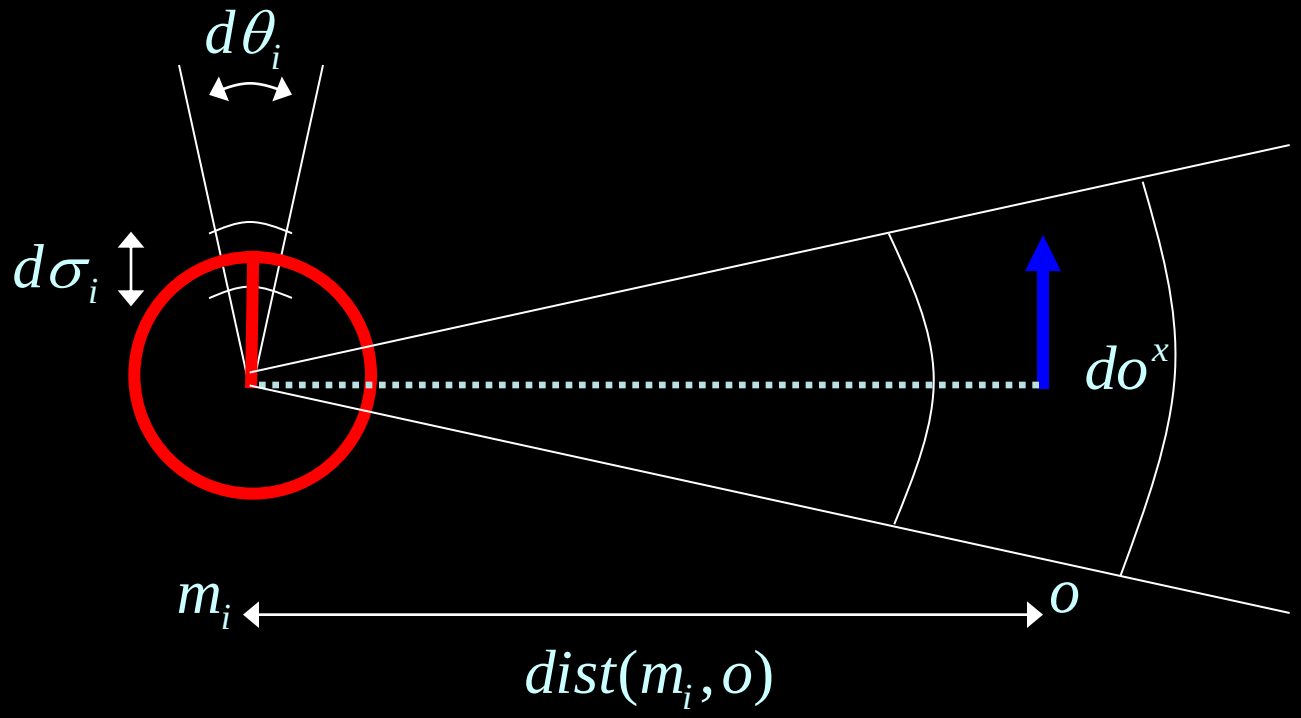
Ambiguity Difficult to Resolve

Choice of OCI

- Is nose-to-forehead OCI optimal?

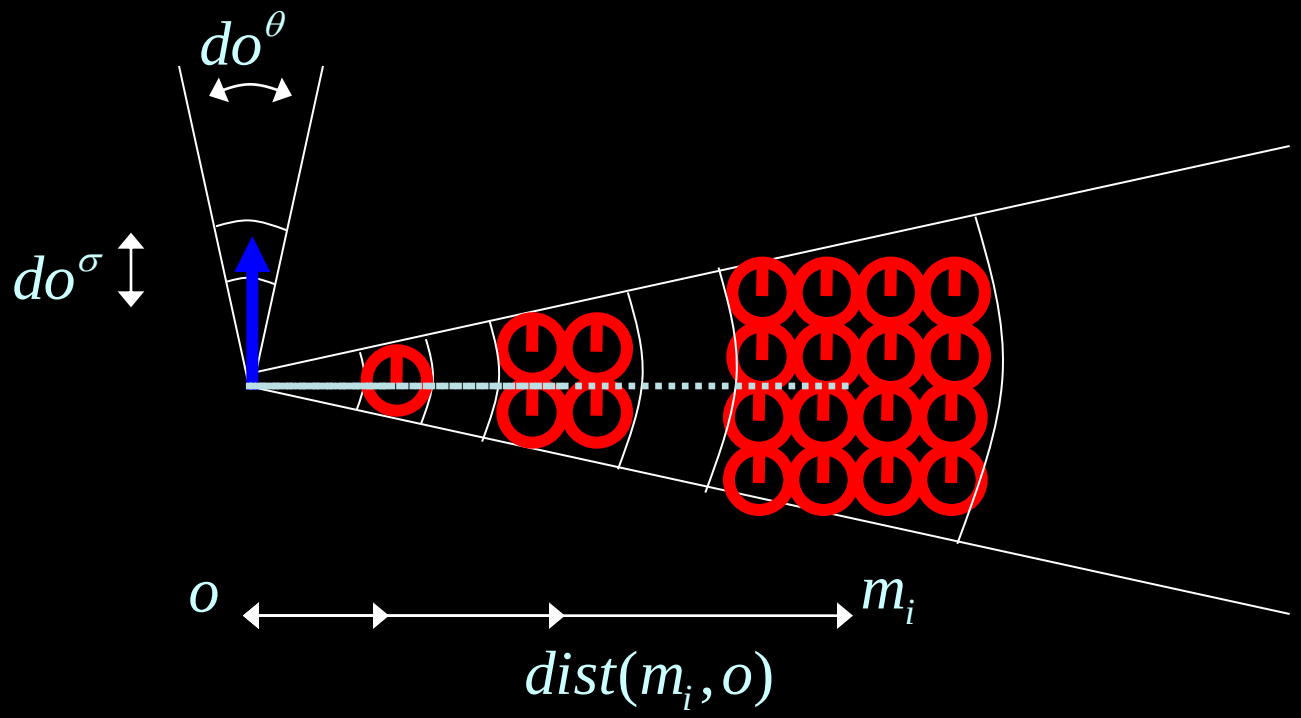
An Optimal OCI?

- Distance between feature and OCI
  - Related to OCI localization error do^x



An Optimal OCI?

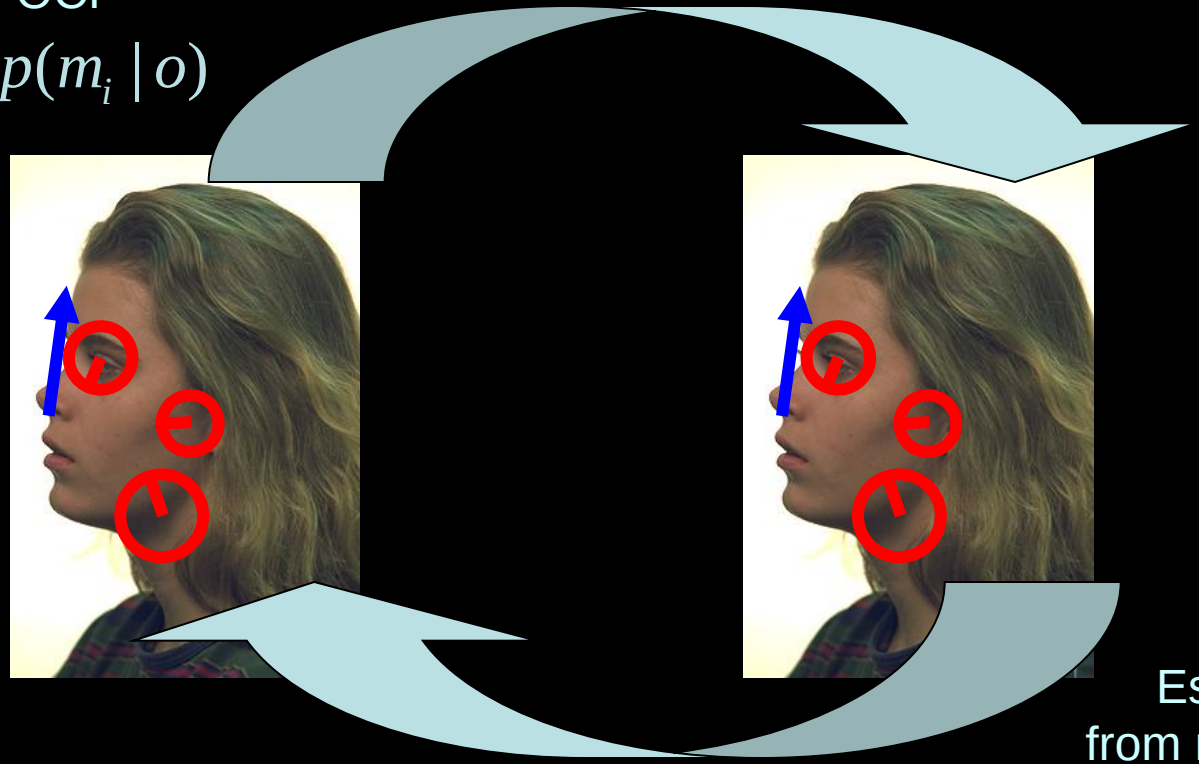
- Distance between feature and OCI
  - Related to probability of a false match



Data-driven OCI Estimation

Learn model features from OCI

$$m_i \leftarrow p(m_i \mid o)$$

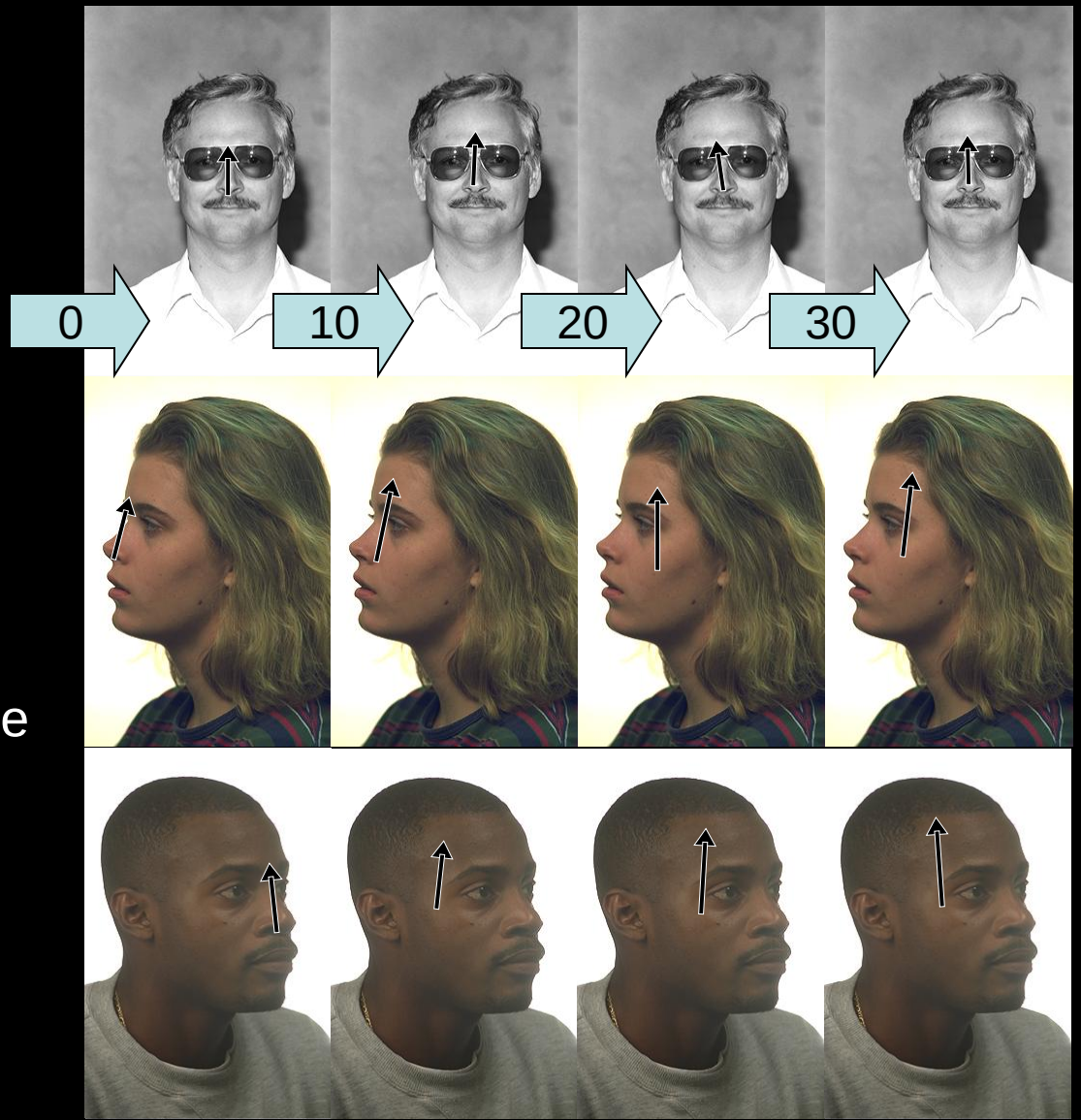


Estimate OCI from model features

$$o \leftarrow \underbrace{argmax}_{o} \left\{ \frac{p(o \mid \{m_i\})}{p(\overline{o} \mid \{m_i\})} \right\}$$

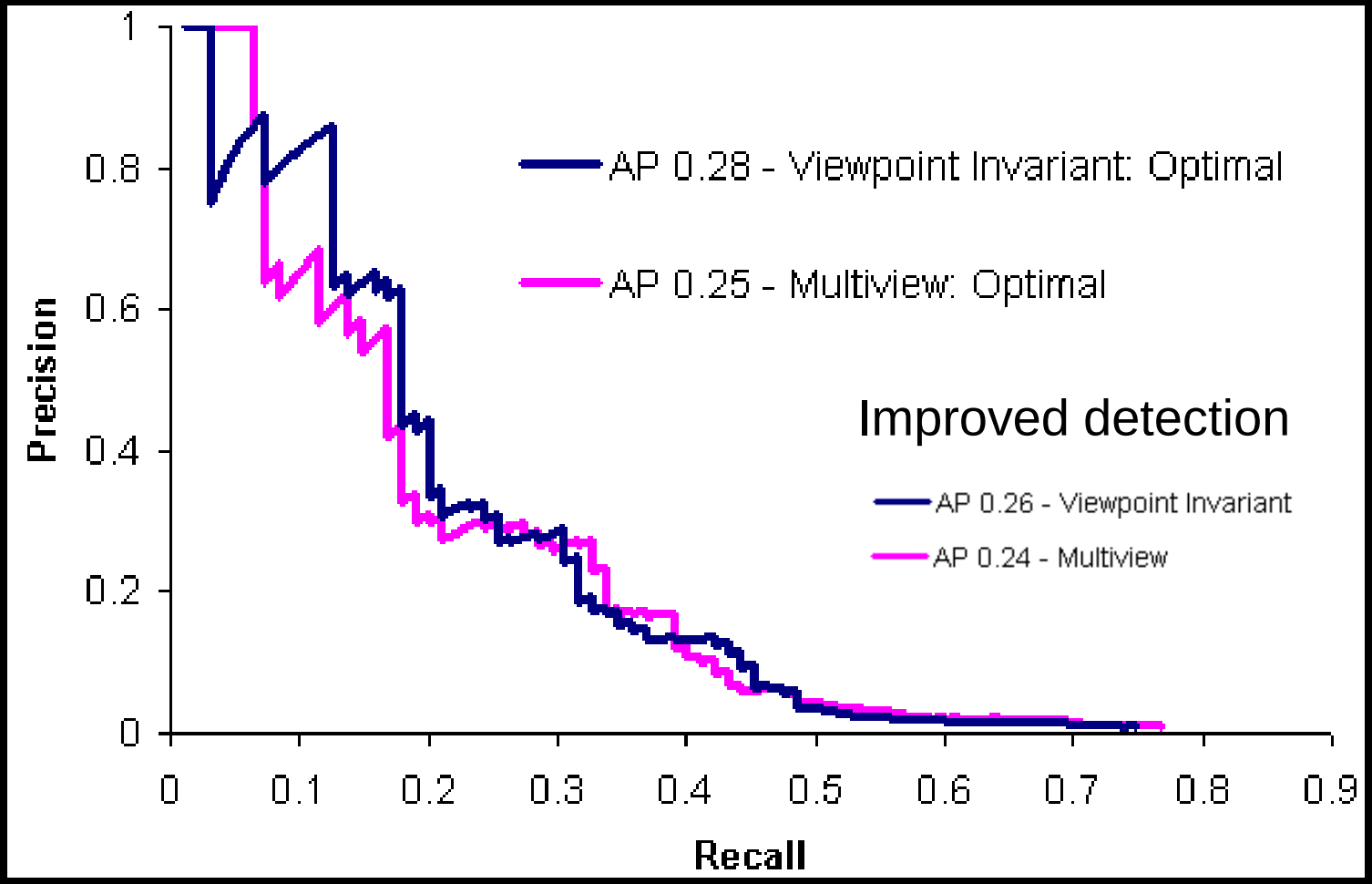
Data-driven OCI Estimation

30 Iterations



OCI remains consistent with 3D head in different views, different people

Comparison


AP 0.28 - Viewpoint Invariant: Optimal
AP 0.25 - Multiview: Optimal
Improved detection
AP 0.26 - Viewpoint Invariant
AP 0.24 - Multiview
Precision
Recall
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

McGill
Centre for
Intelligent Machines

Summary

- Viewpoint-invariant modeling
  - Features related directly to object class
  - Viewpoint information not required
  - Reduces localization ambiguity
- Data-driven OCI
  - Stable, minimizes OCI localization error
  - Consistent with 3D geometry of underlying object class