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

ICCV 3DRR 2007

Overview

- Visual object classes
- Related work
  - Invariant features
  - Probabilistic modeling
- Modeling viewpoint
  - Multi-view
  - Viewpoint-invariant
  - An optimal viewpoint-invariant model

Visual Object Classes

- An object class
  - A set of visually similar objects
  - e.g. cars, faces,...

Visual Object Classes

- Problem
  - Learn appearance from natural imagery
  - Detect and localize new instances

Challenges

- Nuisance parameters
  - Illumination changes
  - In-plane geometrical deformations
  - Partial occlusion
  - Intra-class variation
  - In-depth geometrical deformations (viewpoint variation)
- Generalization
- Computational efficiency

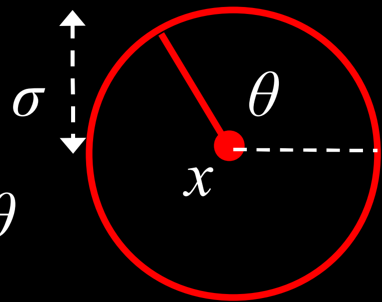
Related Work

- Local invariant image features
  - Salient image characteristics
- Probabilistic modeling
  - Describe appearance in terms image features and probability theory

Local Invariant Image Features

Geometry

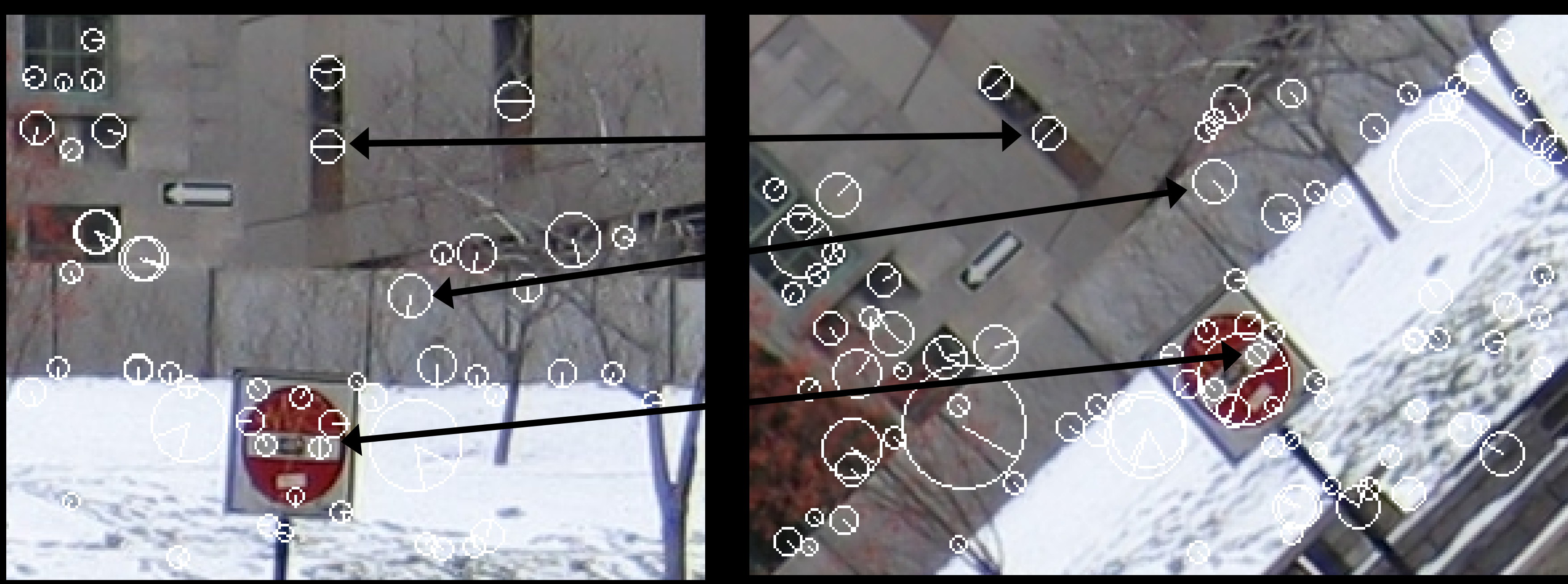
- Location x
- Orientation θ
- Scale σ



Appearance

- Image intensity information
- I.e. Pixels, edges

SIFT: Scale-invariant Feature Transform, Lowe 2004



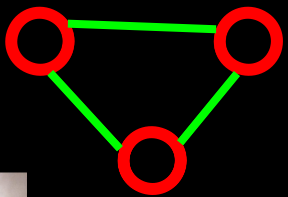
Local Invariant Image Features

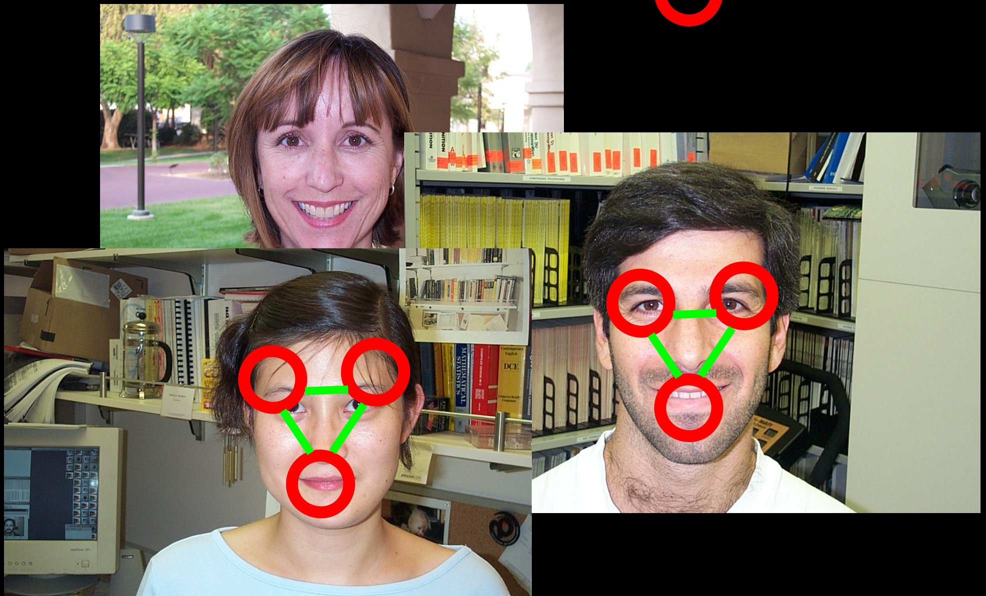
- Address:
  - Nuisance parameters
    - Illumination, in-plane geometrical variation, occlusion
  - Efficiency
  - Generalization
- Do not address:
  - Intra-class variation
  - Viewpoint variation

Probabilistic Models

- Relate local features within a <u>geometrical reference frame</u>
  - e.g. constellations, bounding boxes…

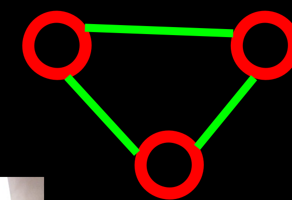
Fergus et al. 2006, …

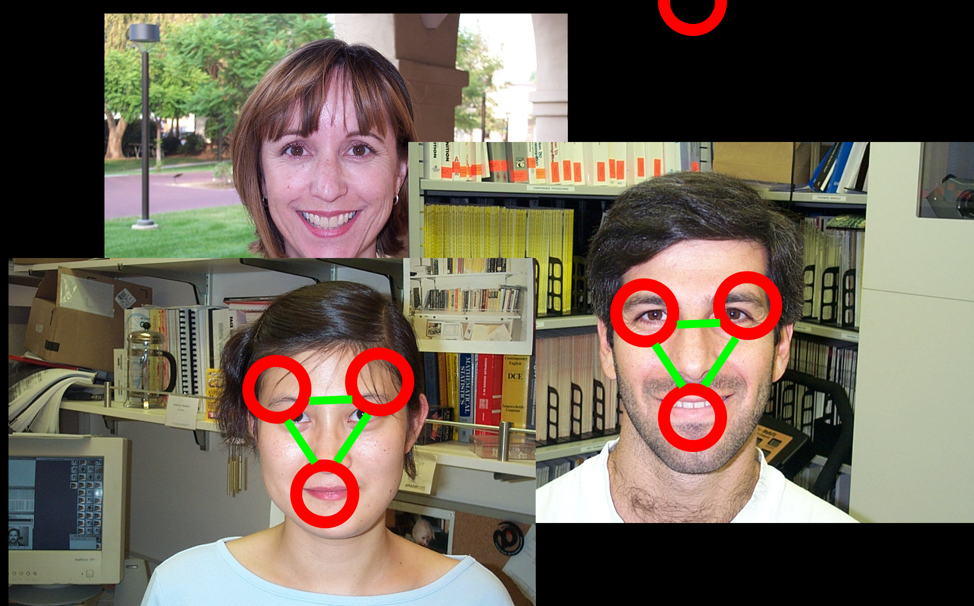




Probabilistic Models

- Address:
  - Intra-class variation
- Challenges:
  - Viewpoint variation





Observation

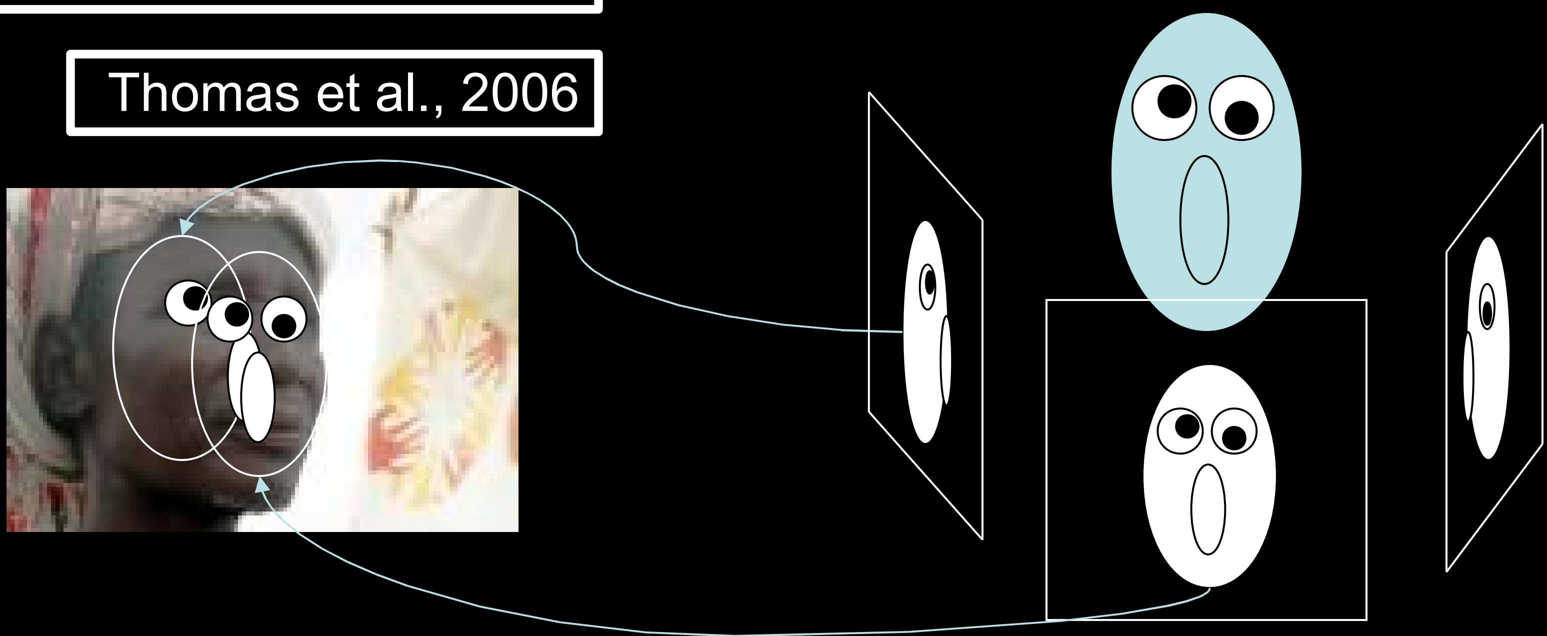
- Features persist over a range of viewpoint.

Probabilistic Models: Multi-view

- Multiple single-view models
- Model viewpoint variable explicitly
- Fit data to nearest view

Koenderink et al., 1979

Thomas et al., 2006



Multi-view: Difficulties

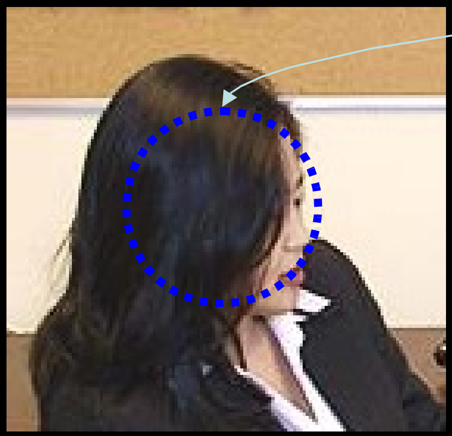
- Viewpoint variable: learning, sampling…

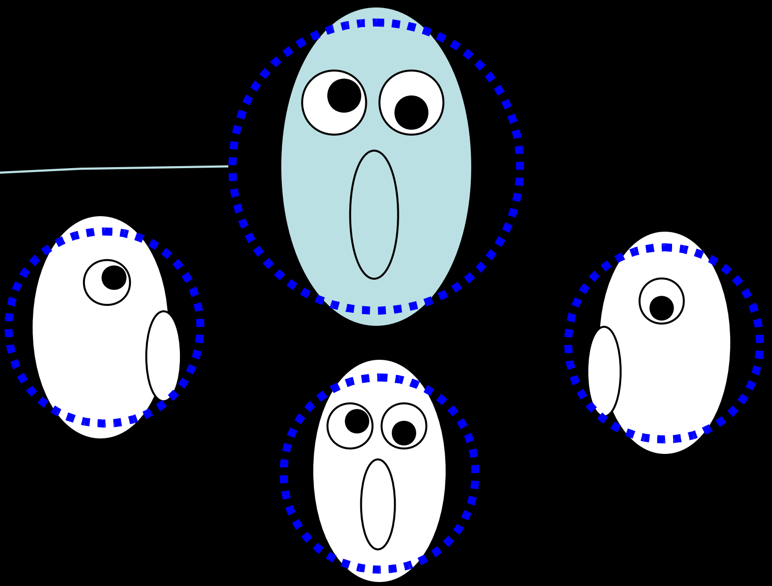
Viewpoint-invariant Model

- Model independent of viewpoint
- Viewpoint invariant reference frame
  - i.e. a perspective invariant
- Infer frame in image

Beiderman, 1987

Toews & Arbel, 2006





Viewpoint-invariance: Advantage

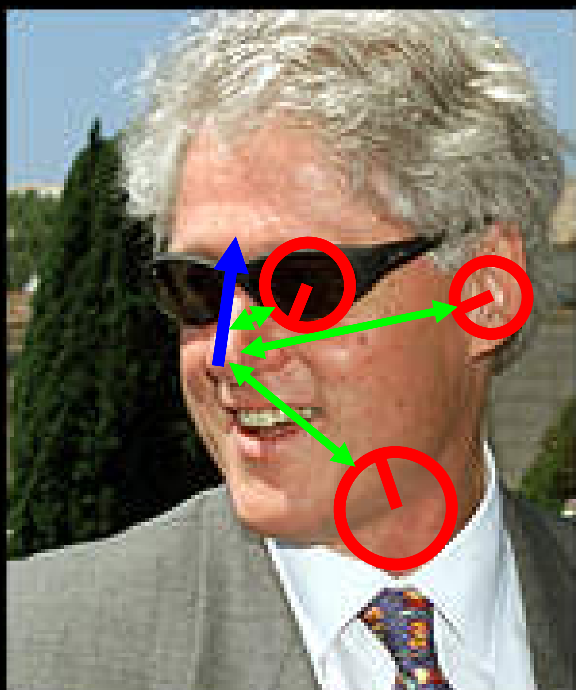
- Simplicity: no viewpoint variable

The Object Class Invariant (OCI)

- A geometrical reference frame that is:
  1) Uniquely defined for each pattern/object class instance.
  2) Invariant to the geometrical transform arising from the imaging process *(perspective projection).*


Toews & Arbel, ICPR 2006

Object Class Invariant Modeling



Object Class Invariant

$o : \{o^b, o^g\}$

Occurrence: binary presence or absence.

Geometry: location, scale, orientation.

Scale-invariant Feature

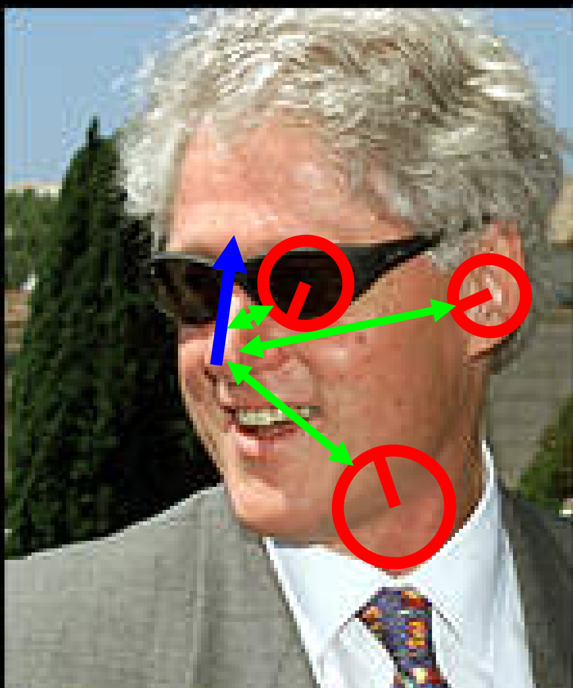
$m_i : \{m_i^b, m_i^g, m_i^a\}$

Appearance: derivative histograms. Note: the OCI is unobservable, and has no appearance!

Transform relating feature and OCI geometries:

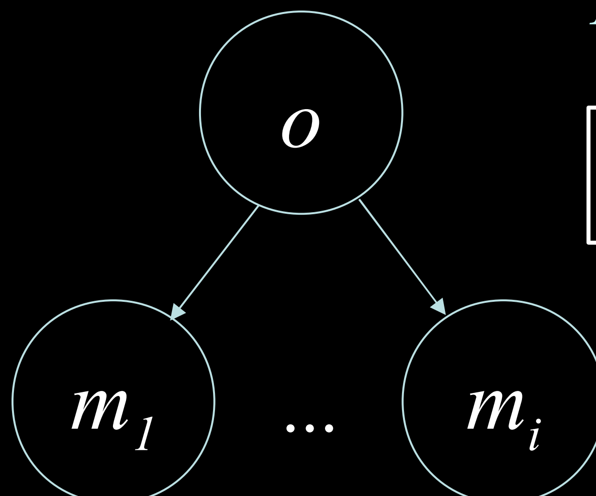
$t_i : m_i^g \rightarrow o^g, o^g = t_i(m_i)$

OCI Model



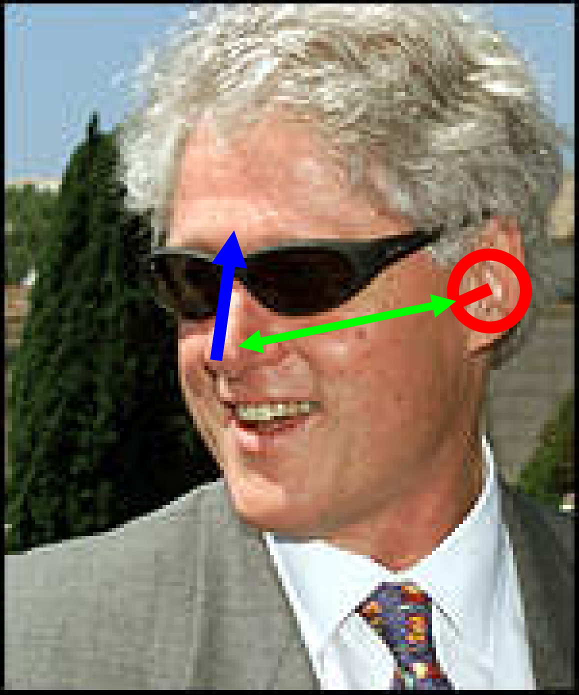
$$p(o \mid \{m_i\}) = \frac{p(o)\, p(\{m_i\} \mid o)}{p(\{m_i\})}$$

$$= \frac{p(o) \prod_i p(m_i \mid o)}{p(\{m_i\})}$$

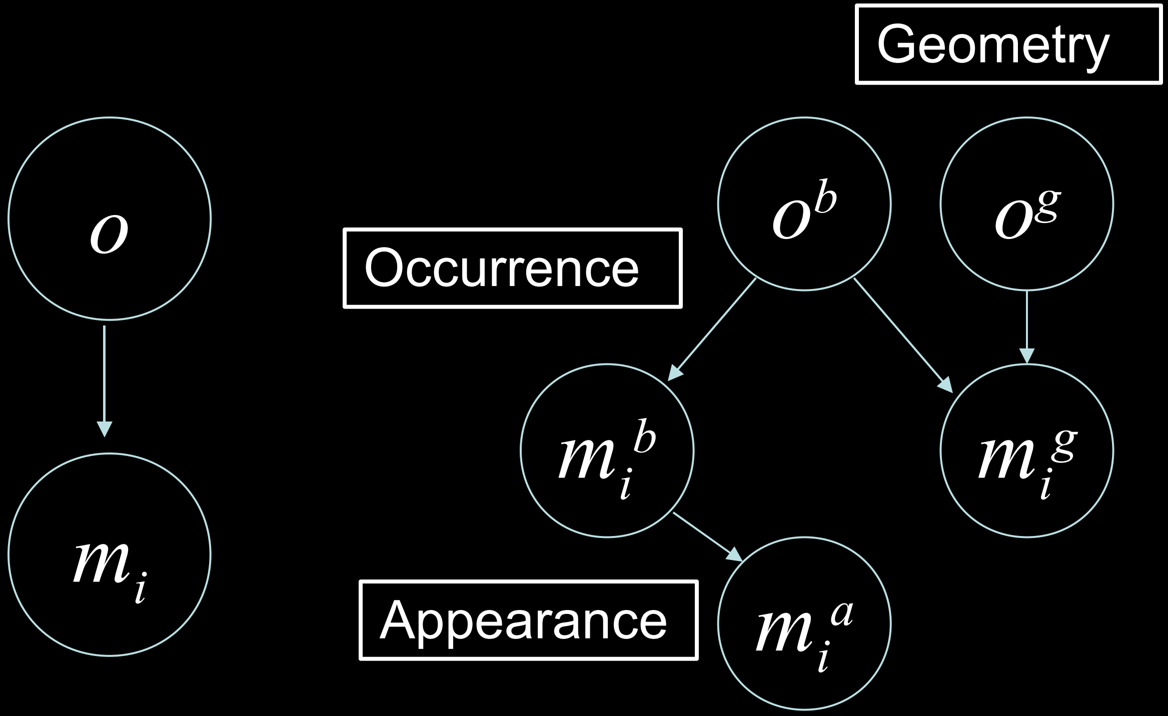


Conditional Feature Independence

OCI Model

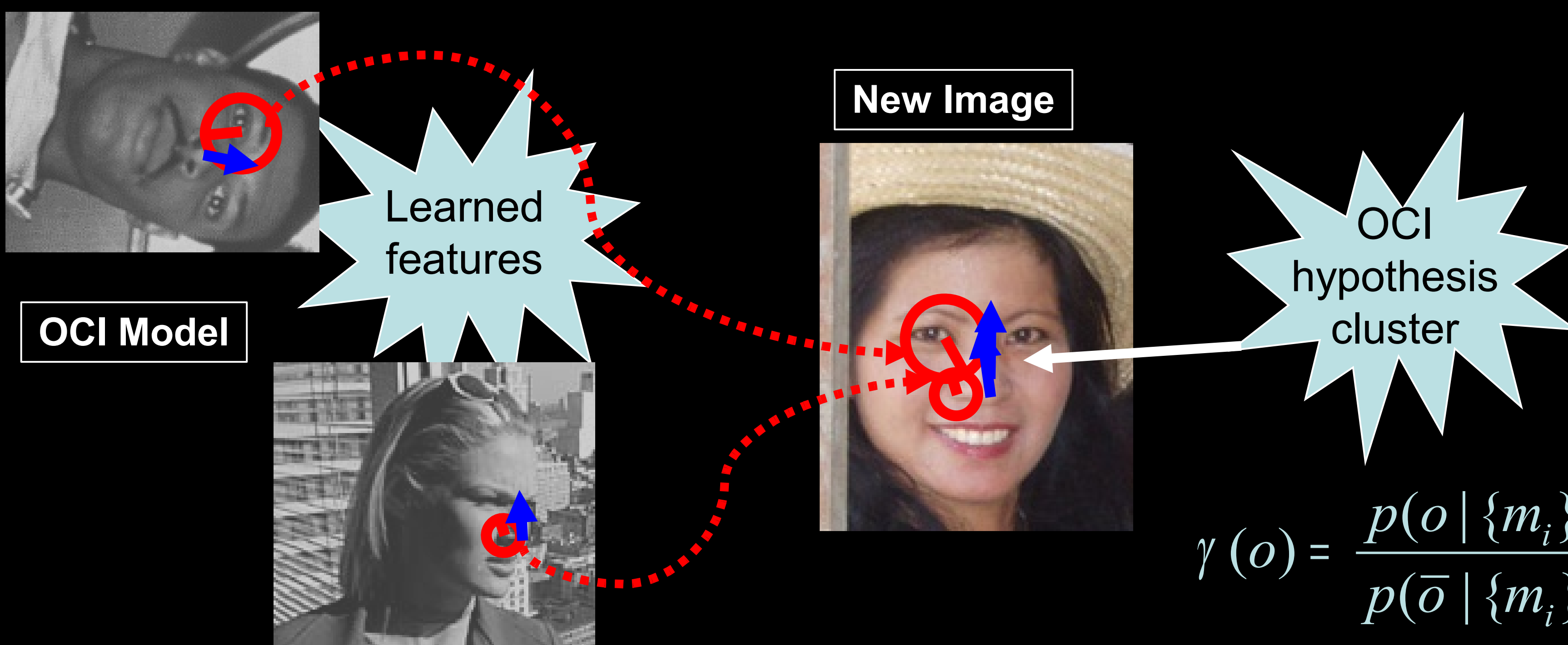


$$p(m_i \mid o) = p(m_i^a \mid m_i^b)\, p(m_i^b \mid o^b)\, p(m_i^g \mid o^b, o^g)$$



OCI Model Fitting

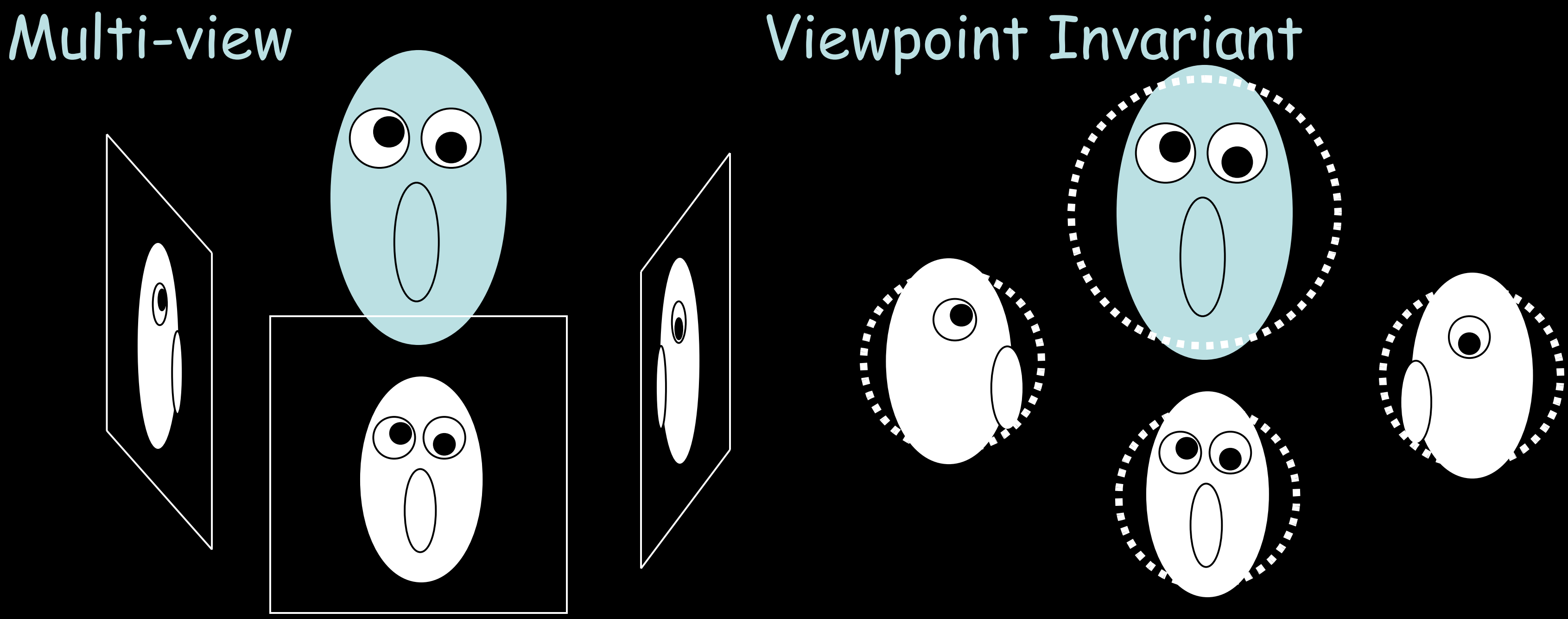
- Identifying OCI instance in new image
  - Probabilistic voting
  - Robust hypothesis clustering



$$\gamma(o) = \frac{p(o \mid \{m_i\})}{p(\bar{o} \mid \{m_i\})}$$

Multi-view or Viewpoint-Invariant?

- Which to use?

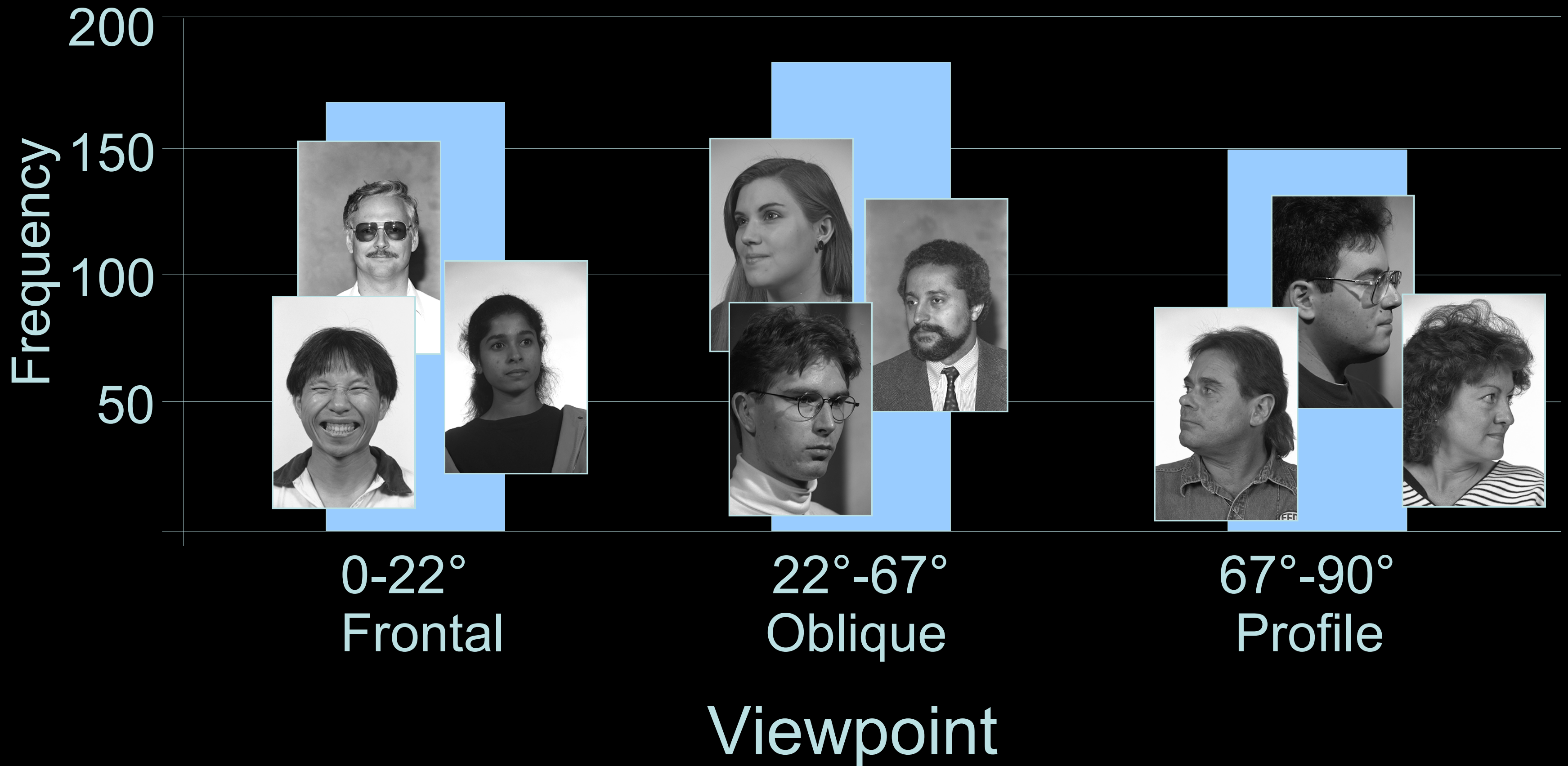


Methodology

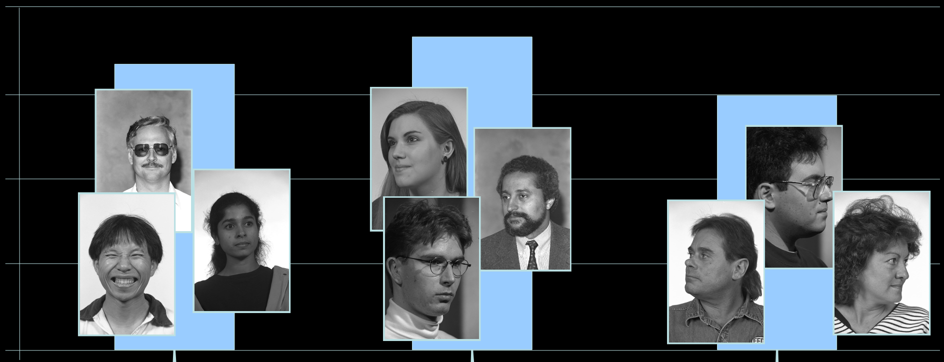
- Learn multi-view and viewpoint-invariant models from the same data.
- Compare detection performance.
- Data: faces, viewpoint variation.

Learning

- Color FERET database
  - 500 unique faces, viewpoint labels

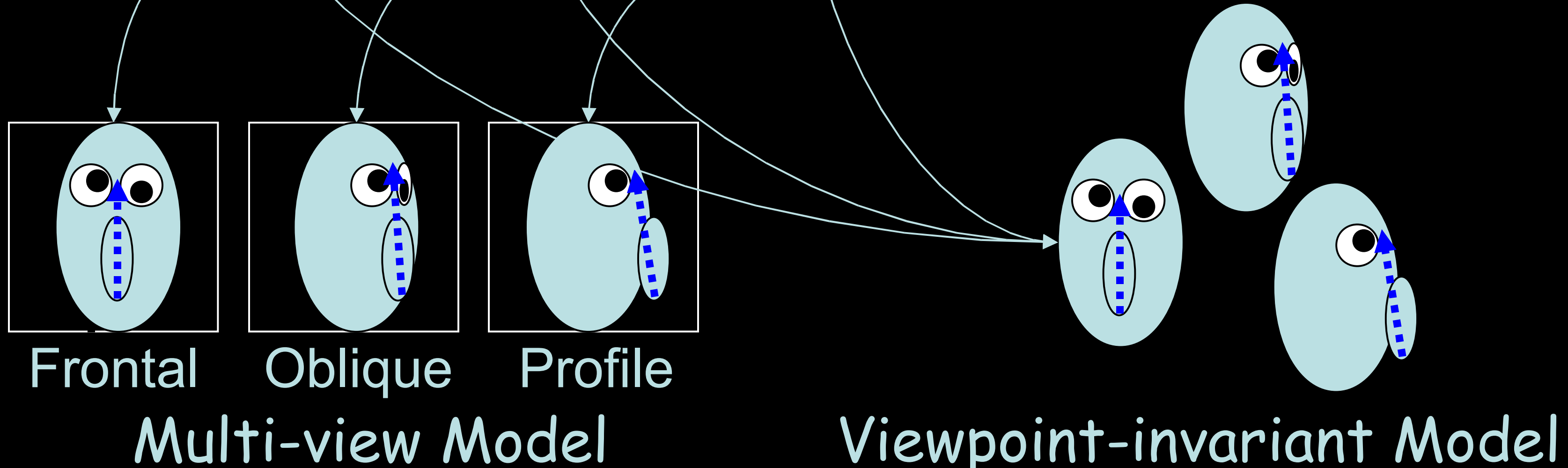


Learning

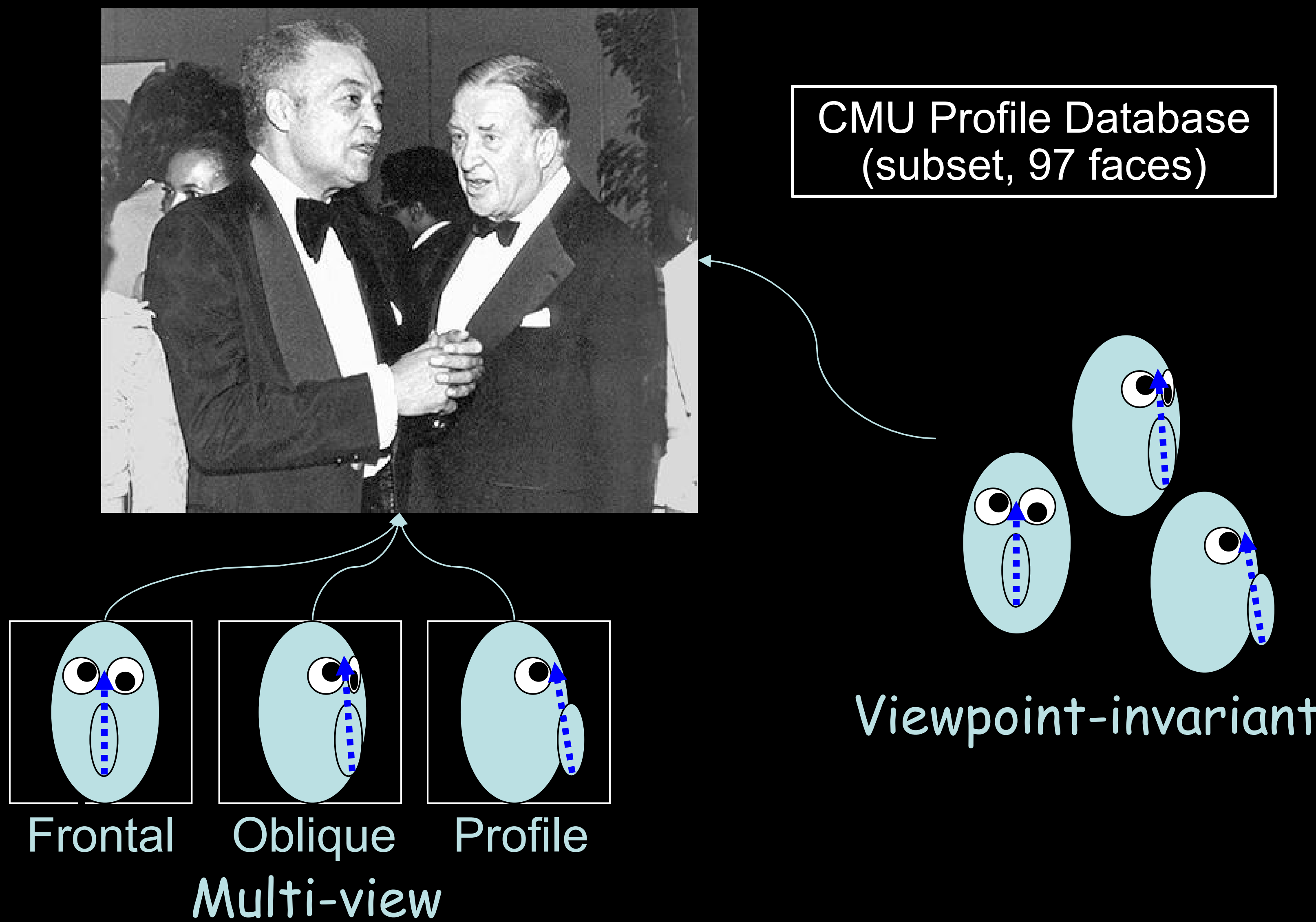


* Manually labelled OCIs

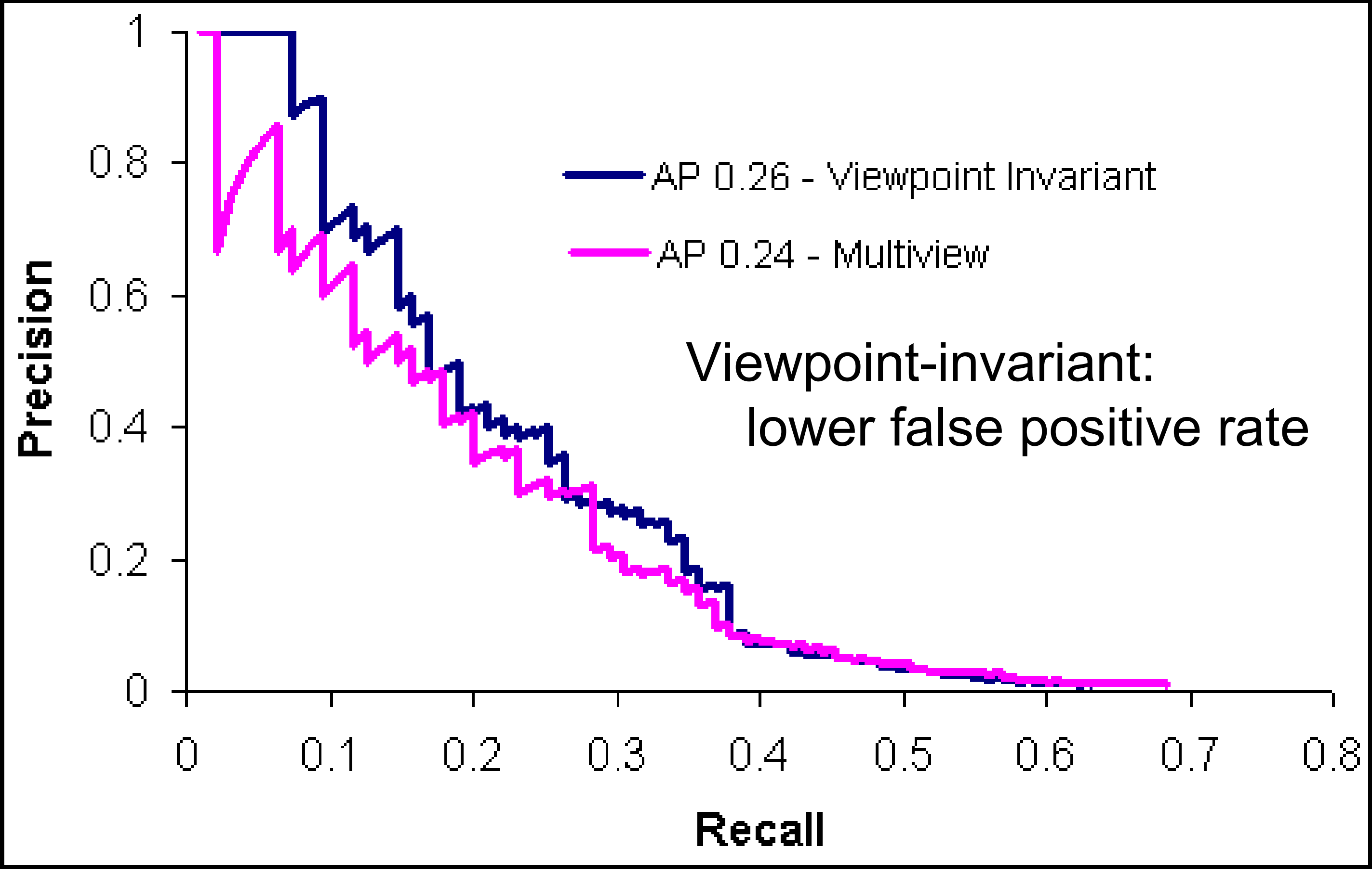
* Robust feature clustering



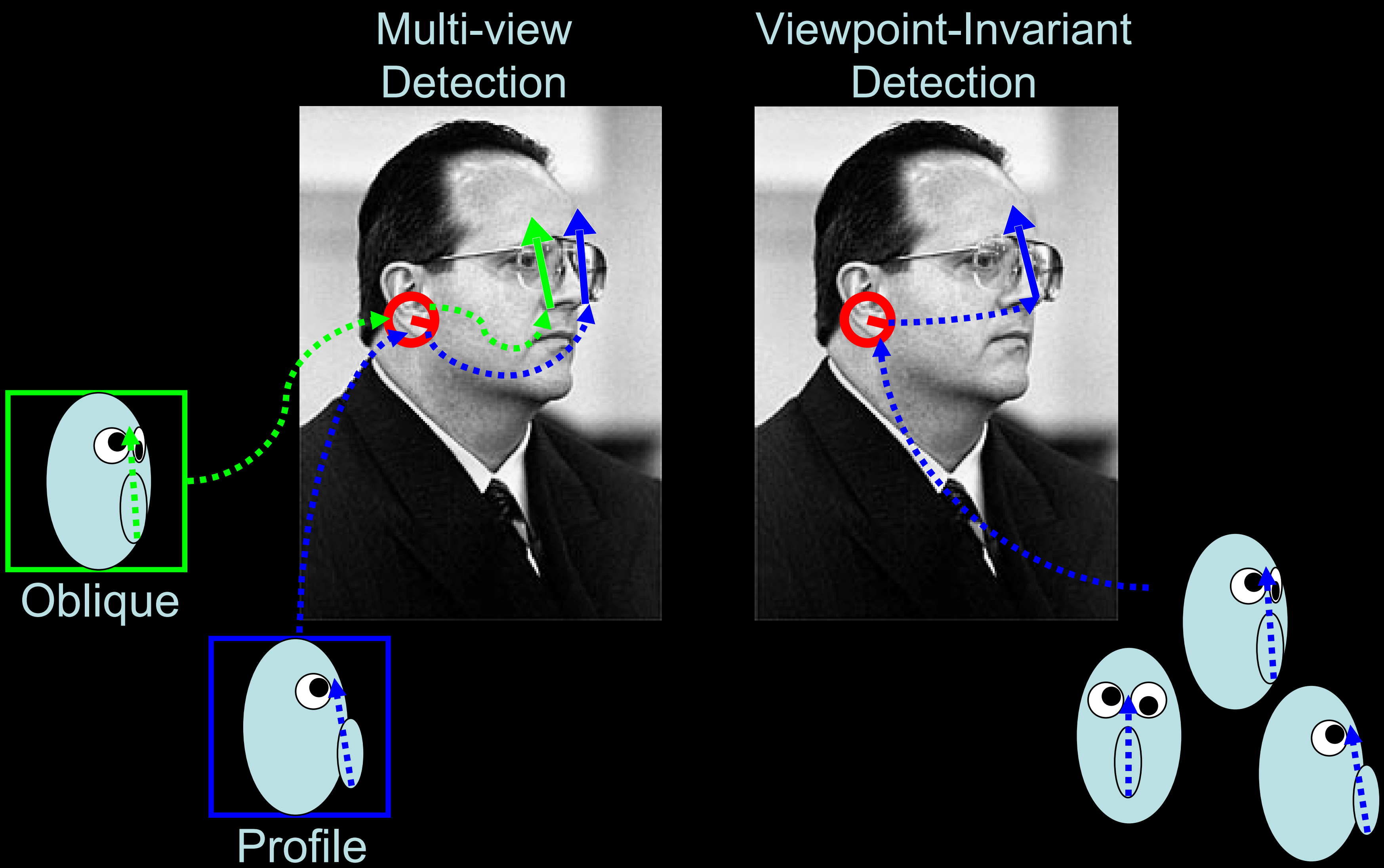
Detection and Localization



Detection Comparison



Example: Localization Ambiguity



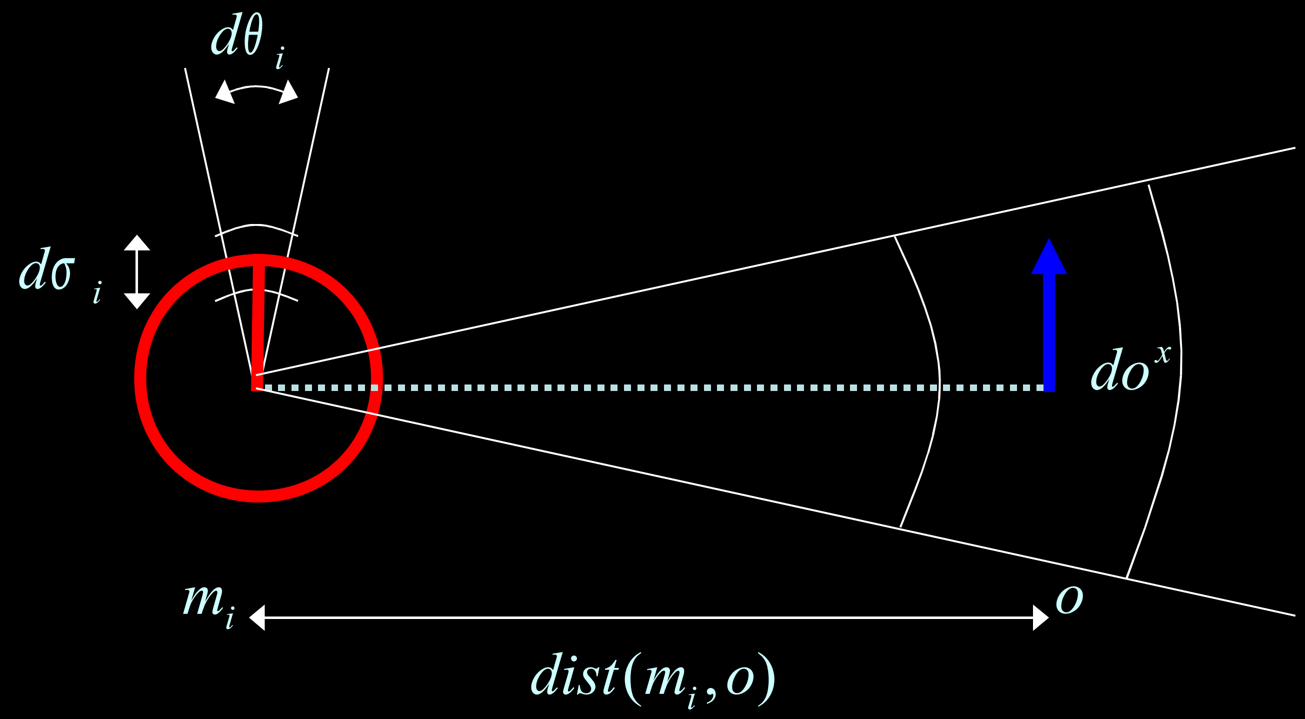
Ambiguity Difficult to Resolve

Choice of OCI

- Is nose-to-forehead OCI optimal?

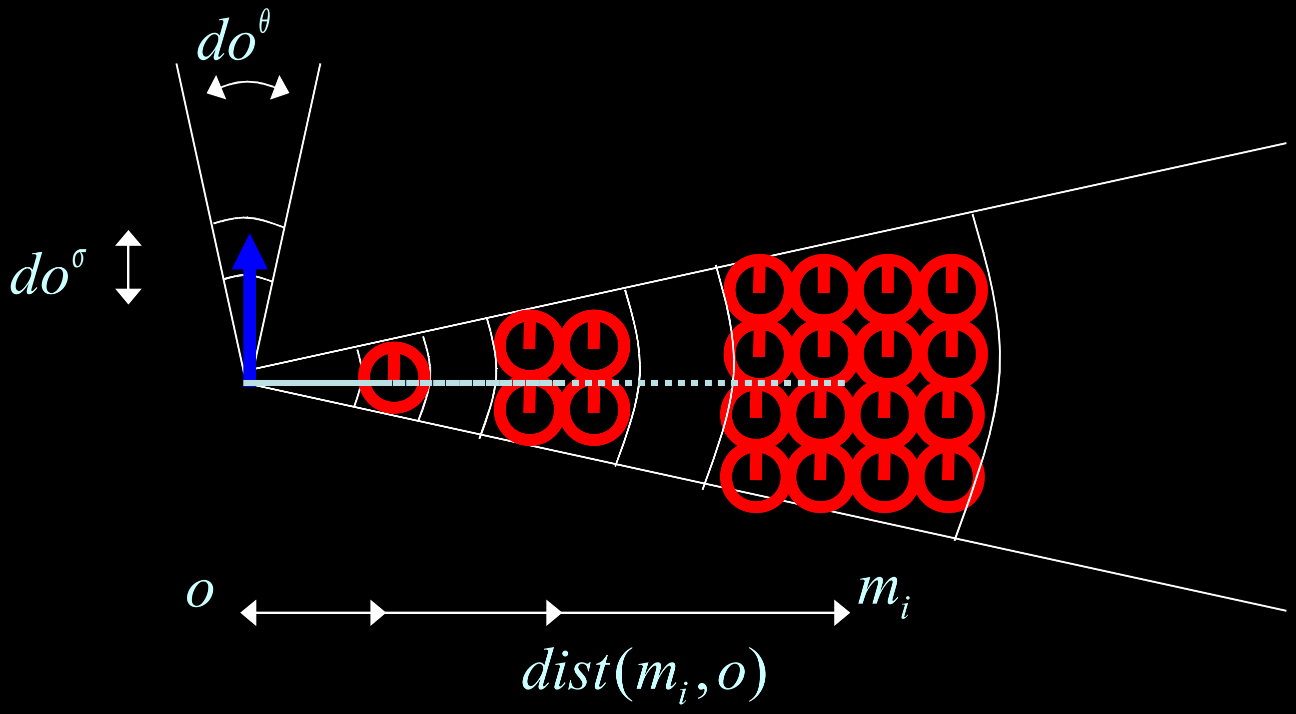
An Optimal OCI?

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



An Optimal OCI?

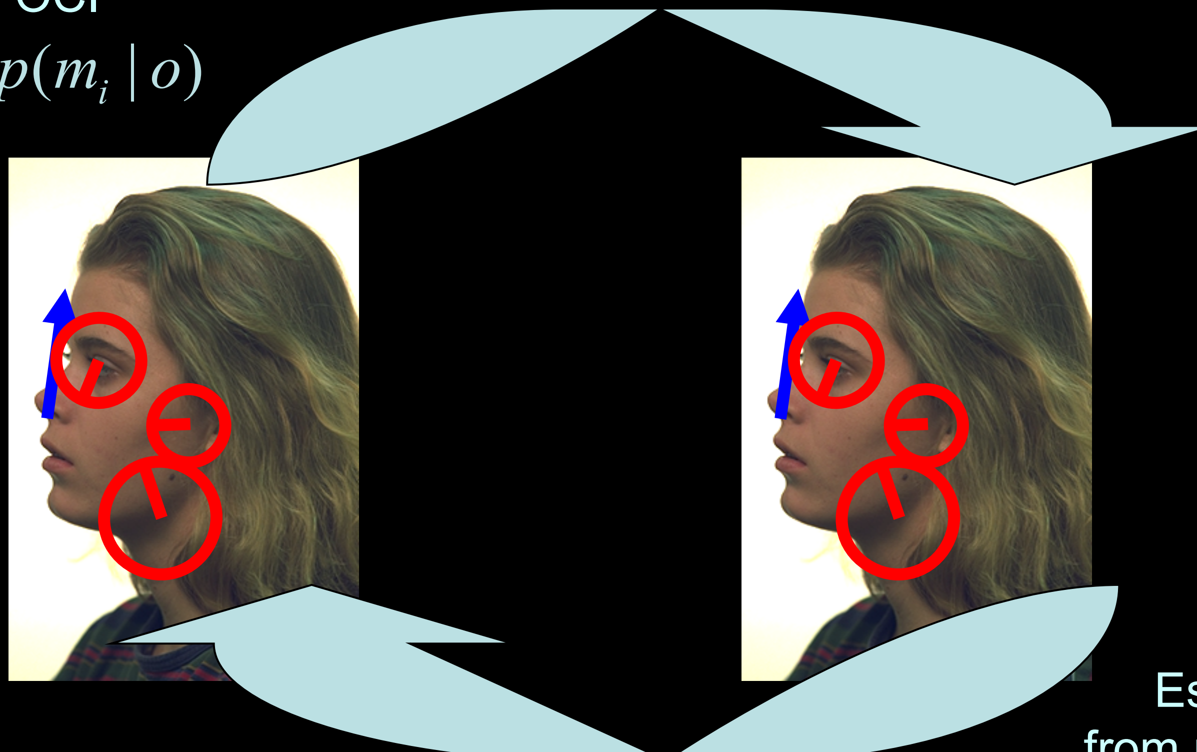
- Distance between feature and OCI
  - Related to probability of a false match



Data-driven OCI Estimation

Learn model features
from OCI

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

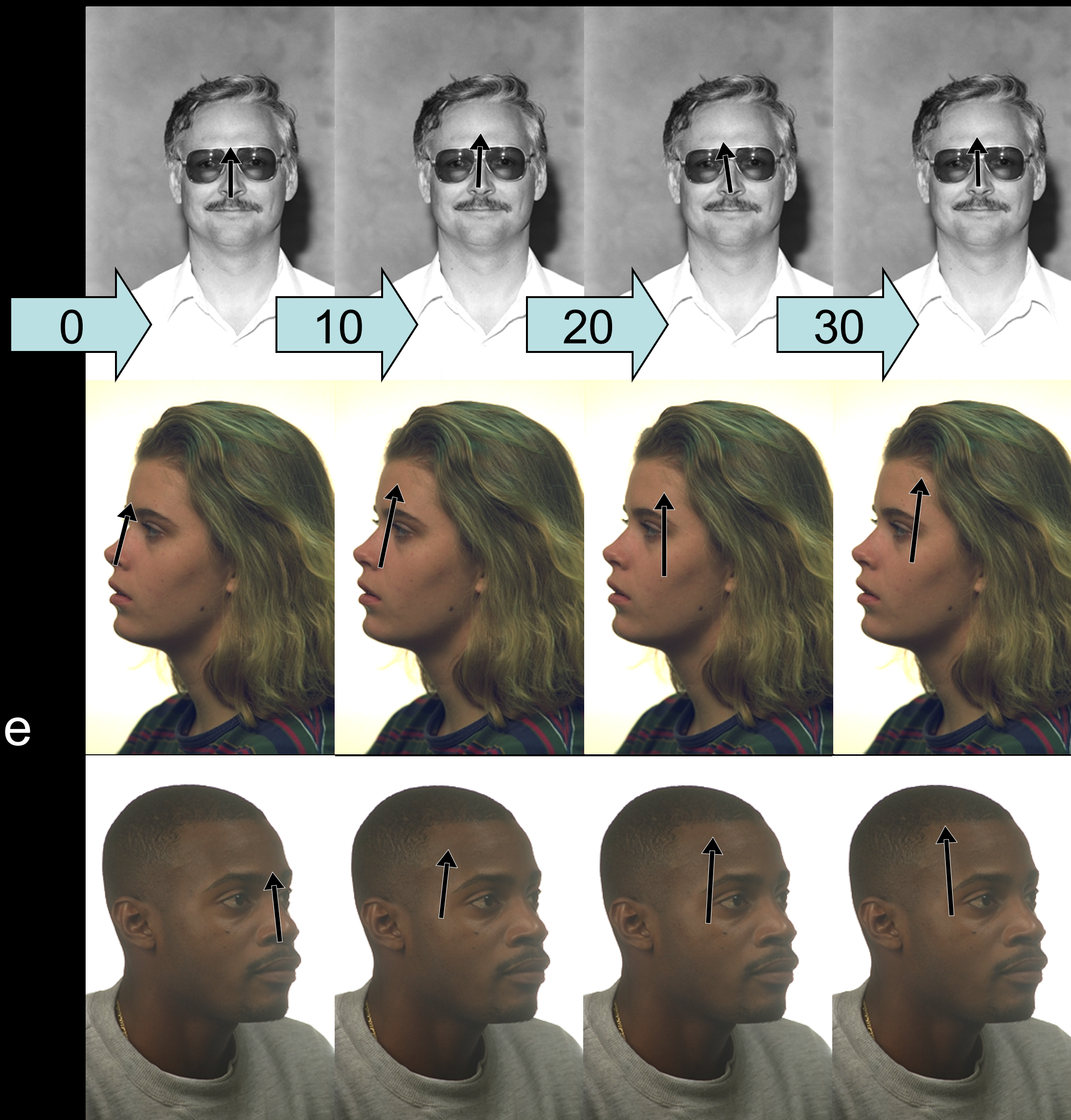


Estimate OCI
from model features

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

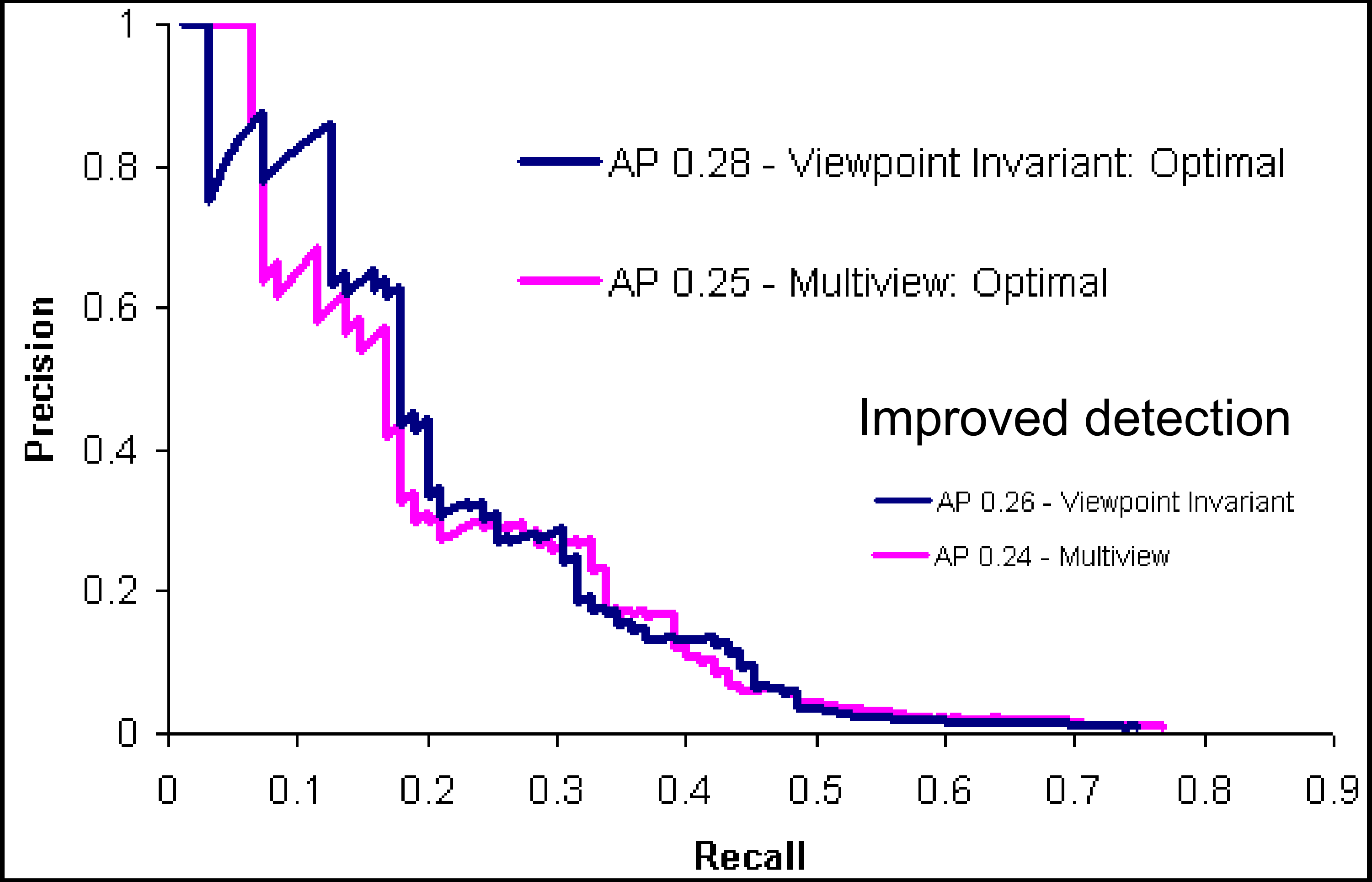
Data-driven OCI Estimation

30 Iterations



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

Comparison



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

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