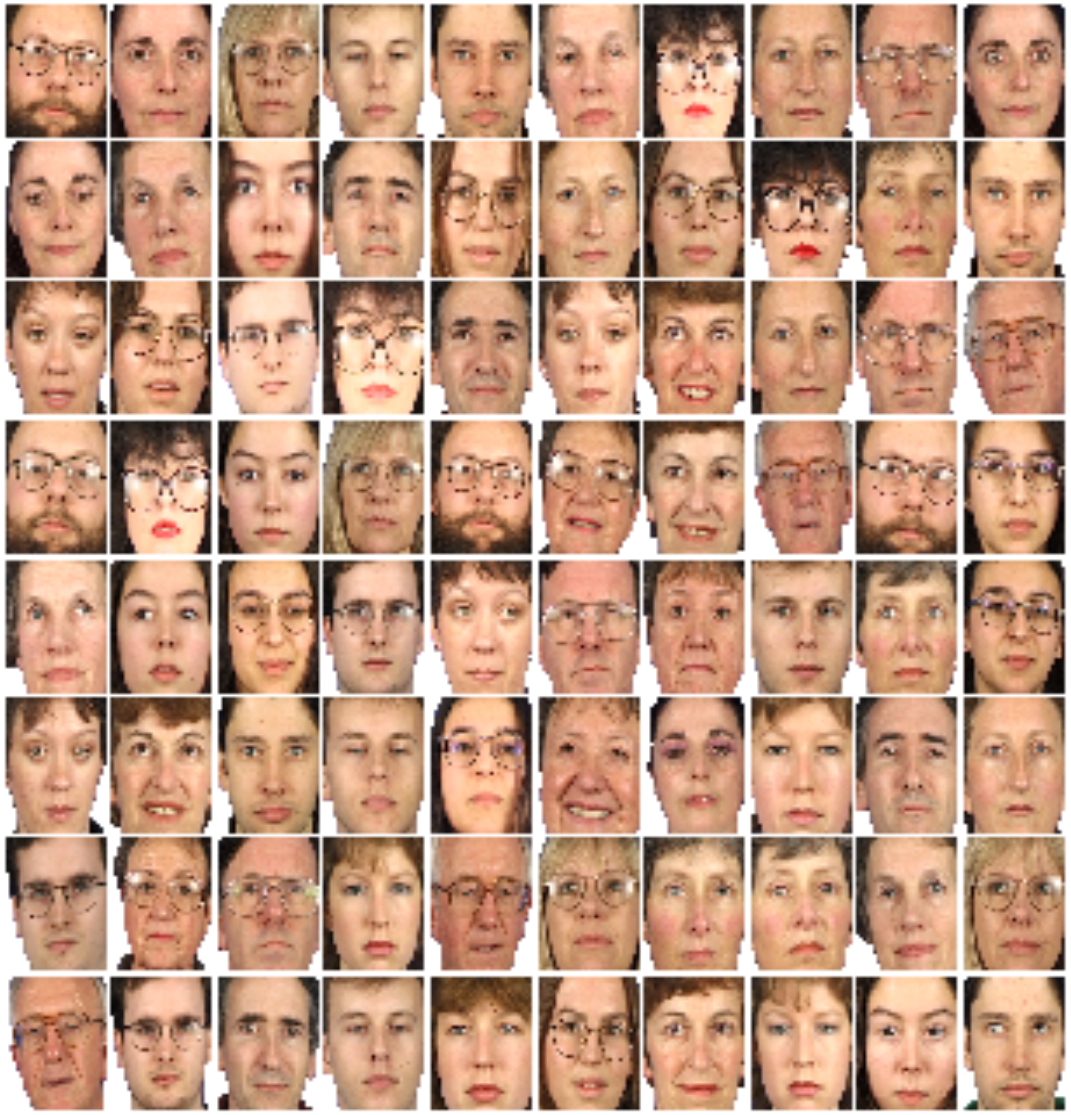
# “Bayesian Identity Clustering”

**Simon J.D. Prince**
Department of Computer Science
University College London.

**James Elder**
Centre for Vision Research
York University.

http://pvl.cs.ucl.ac.uk
s.prince@cs.ucl.ac.uk

# The problem



How many different people are present and which face corresponds to which person?

Could be 80 pics of the same person, or 1 pic each of 80 different people.

Difficult Task:  Compound face recognition + choice of model size

# The problem

Faces don’t necessarily all have the same pose

# Face Clustering

## Application 1: Photo content labelling

System analyzes a set of photos and provides one picture of each different individual present, which you label.

Labels are then propagated to your entire set of photos.


# Face Clustering

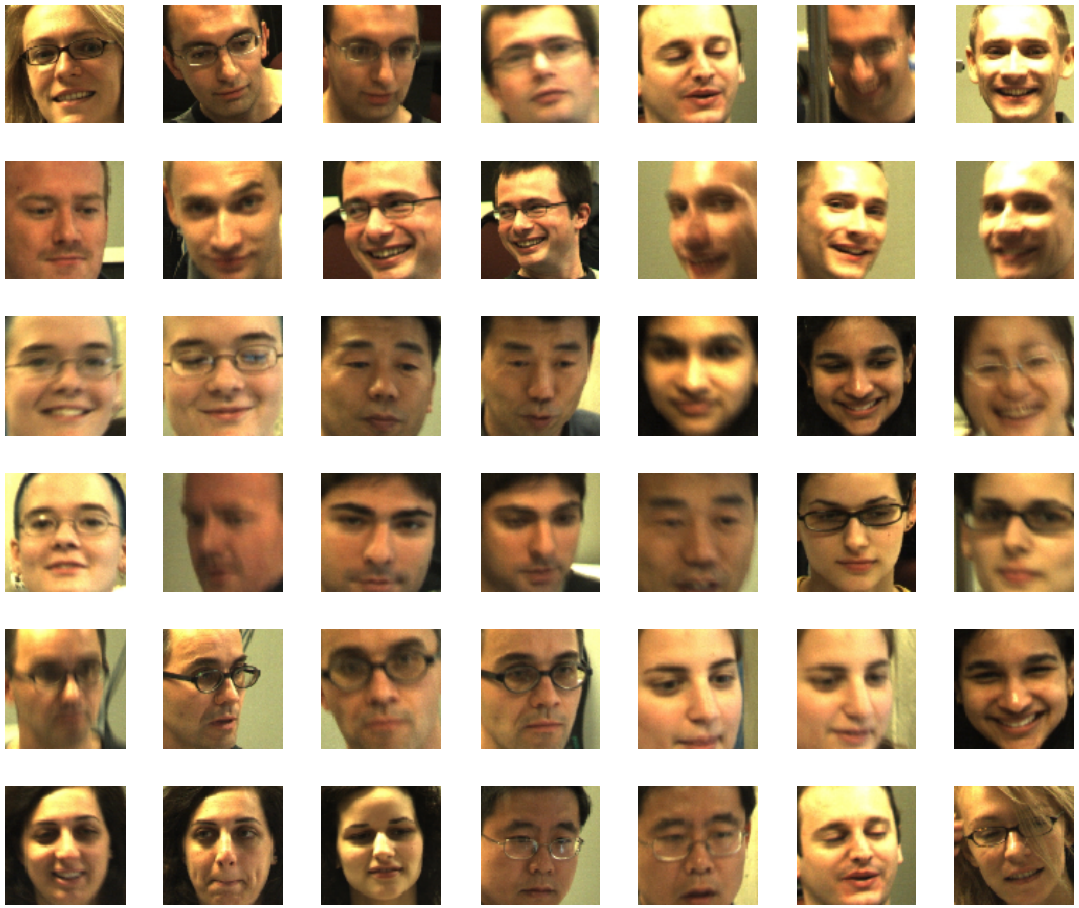
UCL

## Application 2: Security synopses

Here are a collection of face images captured by a pan/tilt security camera.

A crime is committed in the area.

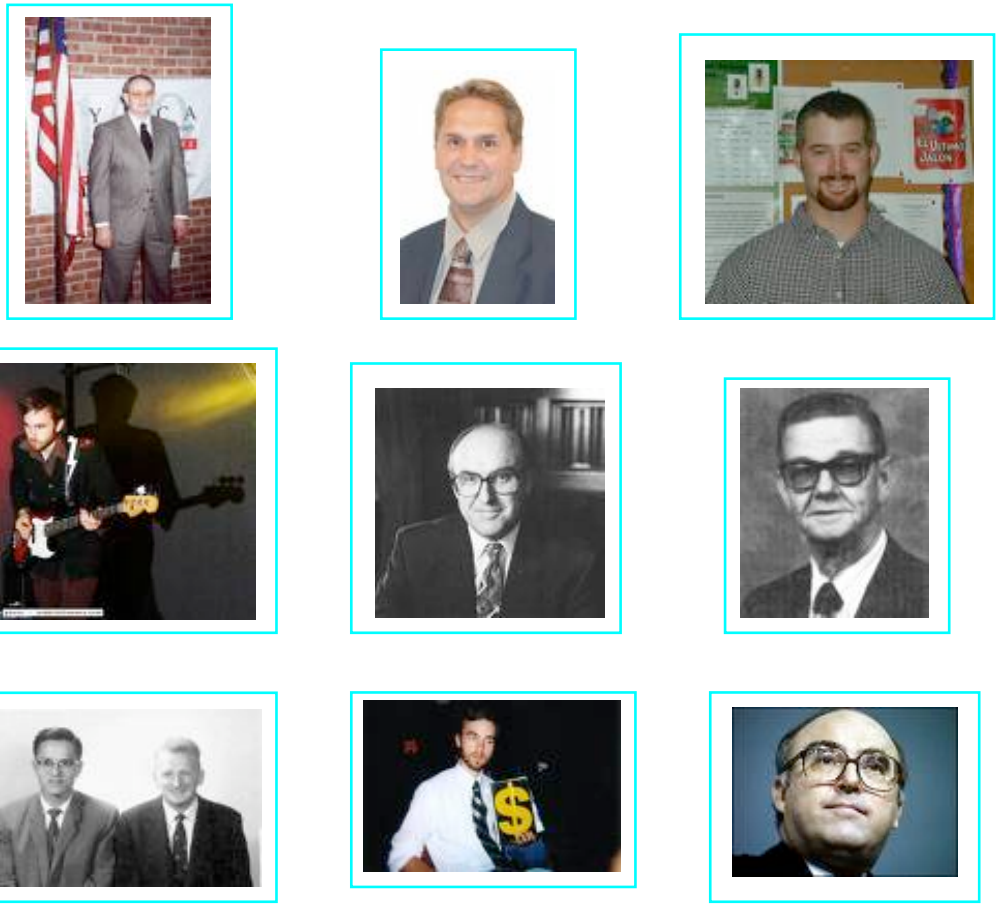
How many people were present in the last few hours, and when did each enter and leave?



## Application 3: Web search

Here are images from Google image search given the query “john smith”.

Identify which pictures belong with which and present only one of each person as options rather than 100’s of pages with repeated images.



# Latent variables for identity

(i) **Face images depend on several interacting factors**: these include the person's identity (signal) and the pose, illumination (nuisance variables).

(ii) **Image generation is noisy**: even in matching conditions, images of the same person differ. This remaining variation comprises unmodeled factors and sensor noise.

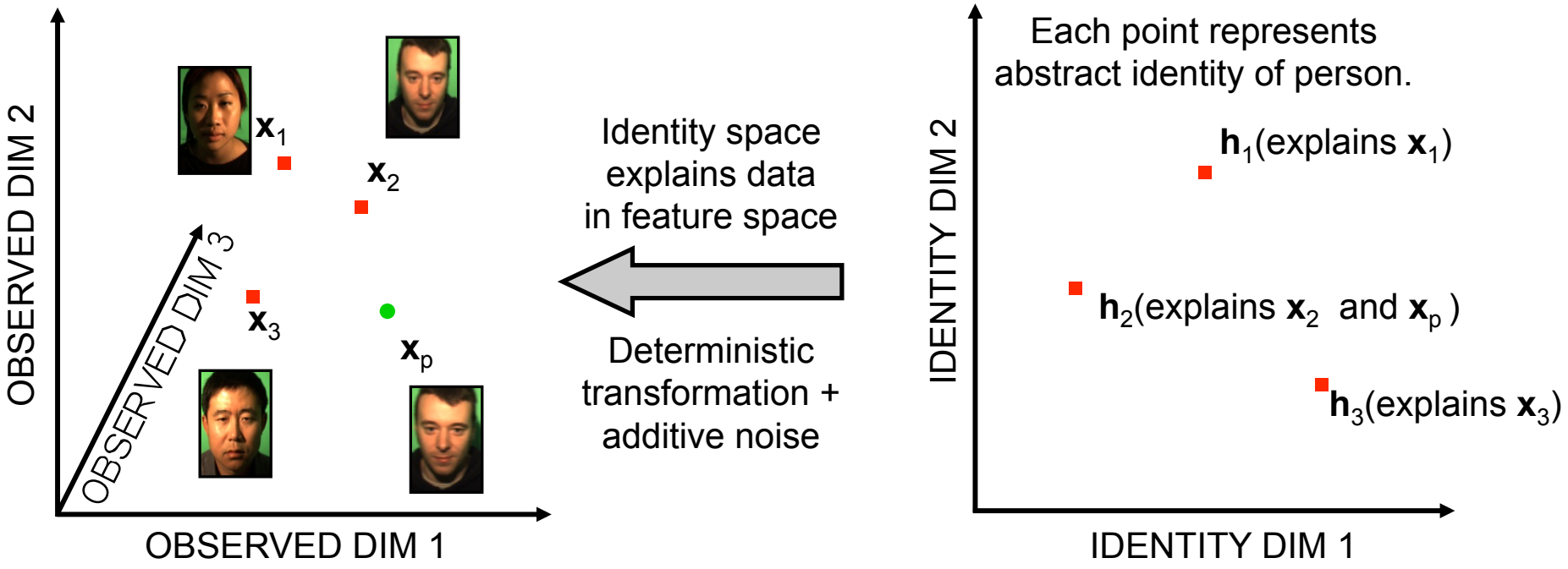
(iii) **Identity can never exactly be known**: since generation is noisy, there will always be uncertainty on any estimate of identity, regardless of how we form this estimate.

(iv) **Recognition tasks do not require identity estimates:** in face recognition, we can ask whether two faces have the *same* identity, regardless of what this identity is.

# Latent Identity Variables

GOAL: To build a generative model of the whole face manifold.

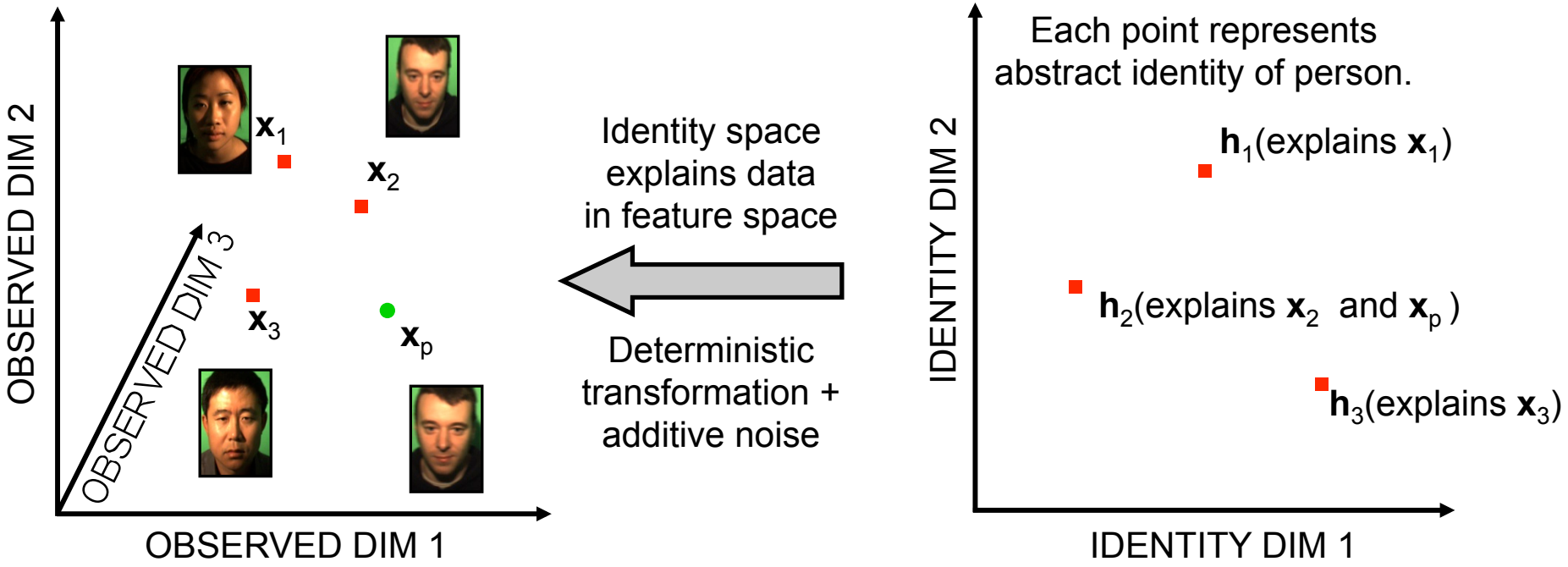
- ✓ We hypothesize an underlying representation of identity (h) and generation process f(.) to create image (a generative model)
- ✓ Add within-individual noise process ε to explain why two images of same person are not identical



# Latent Identity Variables (LIVs)

Hypothesize existence of underlying representation of identity:

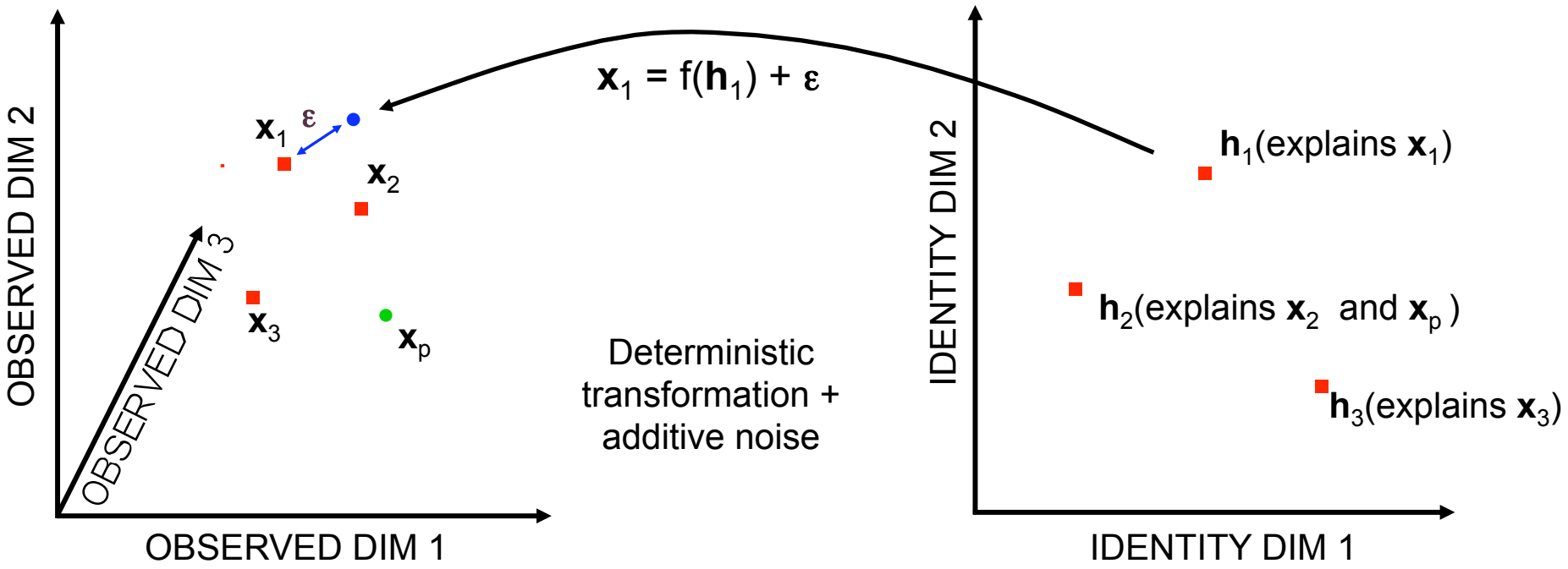
- ✓ If two identity variables take the same value then they represent same person
- ✓ If two identity variables take different values, they represent different people



# Latent Identity Variables (LIVs)

Hypothesize existence of underlying representation of identity:

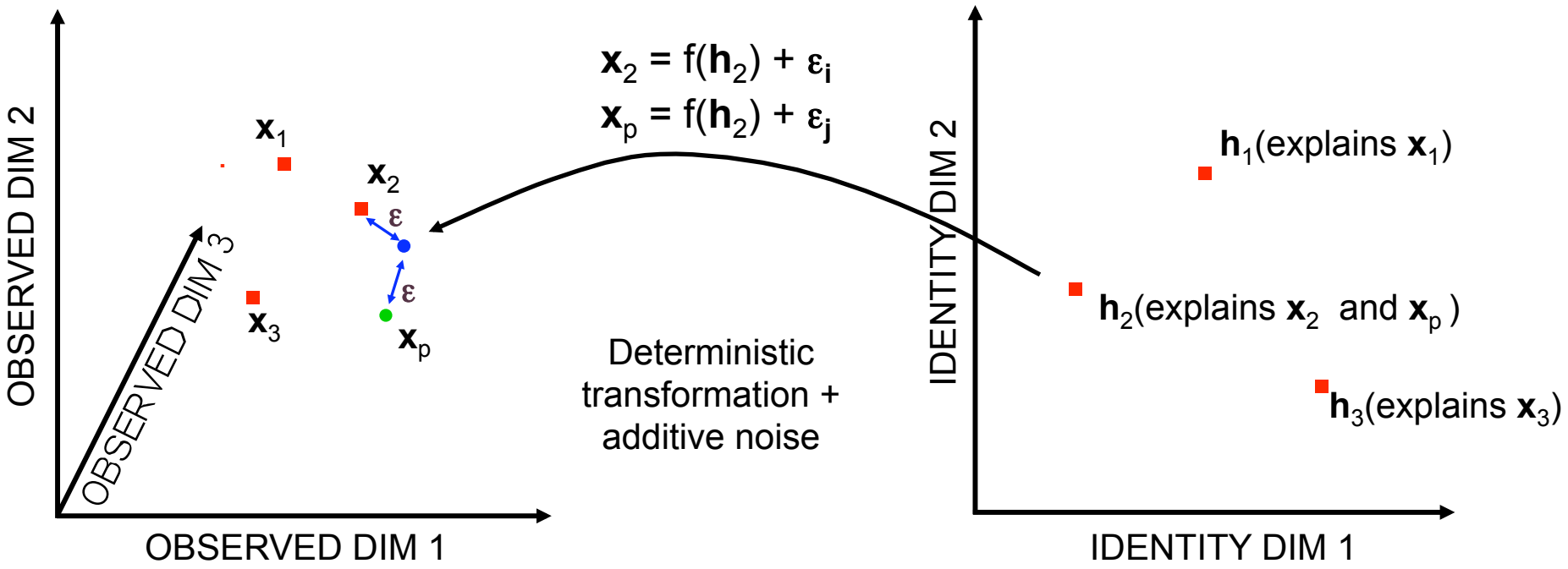
- ✓ If two points in identity space are the same then they represent same person
- ✓ If two points in identity space differ, they represent different people



# Latent Identity Variables (LIVs)

Hypothesize existence of underlying representation of identity:

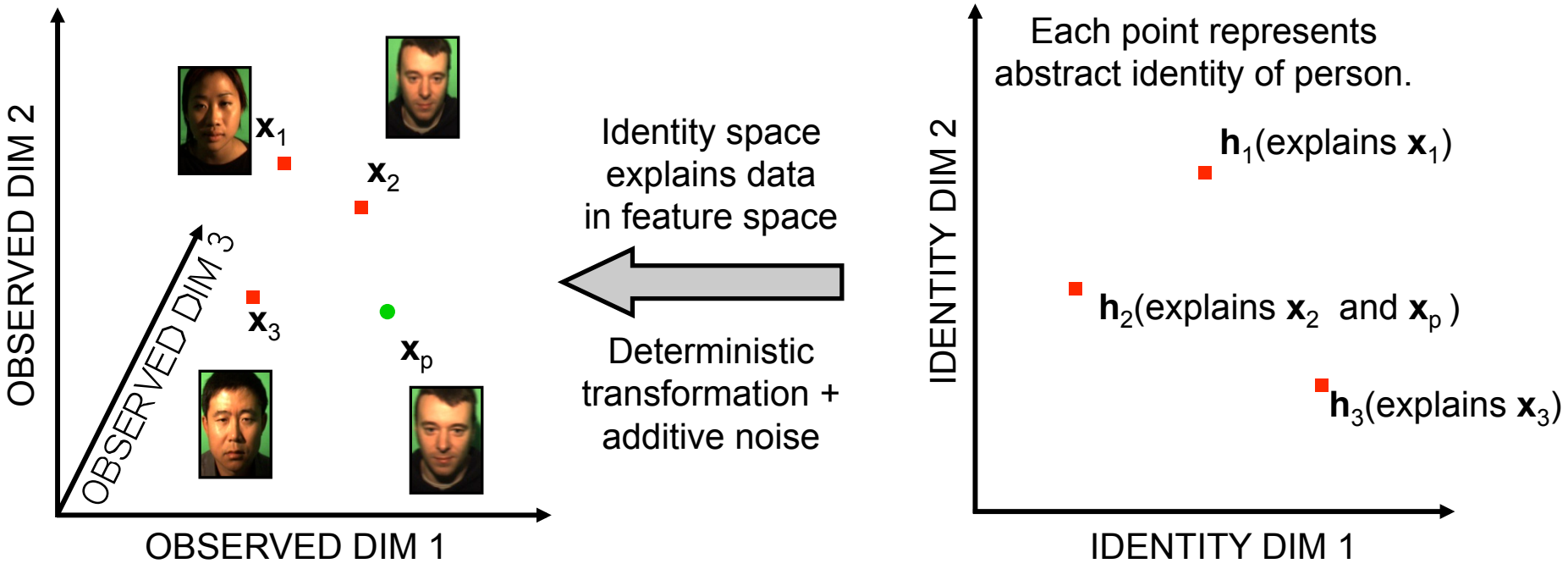
- ✓ If two points in identity space are the same then they represent same person
- ✓ If two points in identity space differ, they represent different people



# Latent Identity Variables (LIVs)

So a LIV model partitions of the data into signal and noise. In recognition we:

- ✓ Compare the probability of data under different assignments between the data (left figure) and identity (right figure)
- ✓ Acknowledge that values of identity variables are fundamentally uncertain so consider all possibilities (i.e. marginalize over them) – never estimate identity!



# Probabilistic Linear Discriminant Analysis (Prince & Elder 2007)

$$\mathbf{x}_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij}$$

Observed data from j th image of i th indivual

Overall mean

Weighted sum of basis functions **F** for between individual variation (identity)

Weighted sum of basis functions **G** for within-individual variation

noise

# Probabilistic LDA

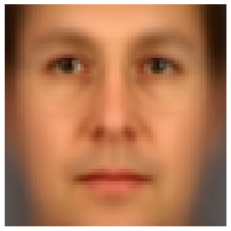
$$\mathbf{x}_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij}$$

Or:

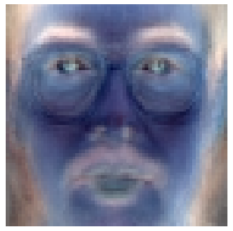
$$\begin{aligned} Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \theta) &= \mathcal{G}_{\mathbf{x}}\left[\mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \Sigma\right] \\ Pr(\mathbf{h}_i) &= \mathcal{G}_{\mathbf{h}}\left[0, \mathbf{I}\right] \\ Pr(\mathbf{w}_{ij}) &= \mathcal{G}_{\mathbf{w}}\left[0, \mathbf{I}\right] \end{aligned}$$


IMAGE, $x_j$

=


MEAN, $\mu$

+ $h_1$


F(:,1)

+ $h_2$

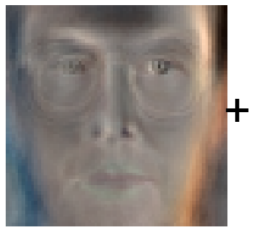

F(:,2)

+ $h_3$

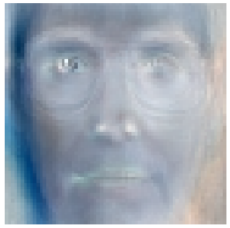

F(:,3)

+

+ $w_{1j}$


G(:,1)

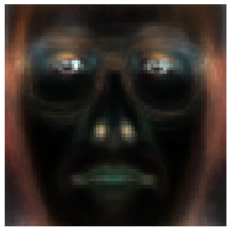
+ $w_{2j}$


G(:,2)

+ $w_{3j}$


G(:,3)

+


NOISE, $\Sigma$

# Within individual and between individual variation

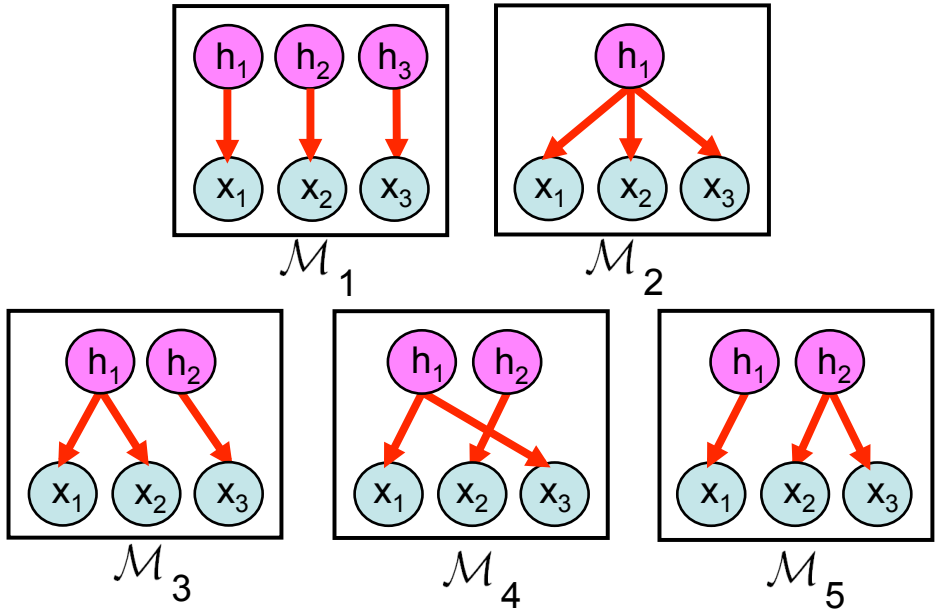
$$\mathbf{x}_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij}$$

Between-individual variation

Within-individual variation


# Model Comparison

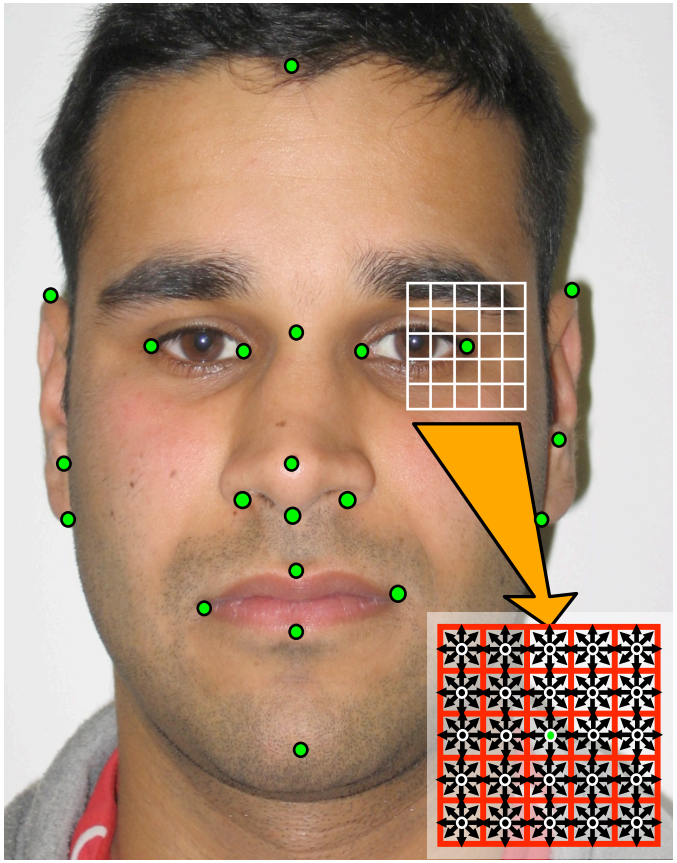
Frame clustering as model comparison




$$Pr(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathcal{M}_1) = \int Pr(\mathbf{x}_1, \mathbf{h}_1) d\mathbf{h}_1 \int Pr(\mathbf{x}_2, \mathbf{h}_2) d\mathbf{h}_2 \int Pr(\mathbf{x}_3, \mathbf{h}_3) d\mathbf{h}_3$$

$$Pr(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathcal{M}_2) = \int Pr(\mathbf{x}_1, \mathbf{h}_1) Pr(\mathbf{x}_2, \mathbf{h}_1) Pr(\mathbf{x}_3, \mathbf{h}_1) d\mathbf{h}_1$$

# Data Extraction



XM2VTS database

Trained with 8 images each of 195 individuals

Test with remaining 100 individuals

Find facial feature points

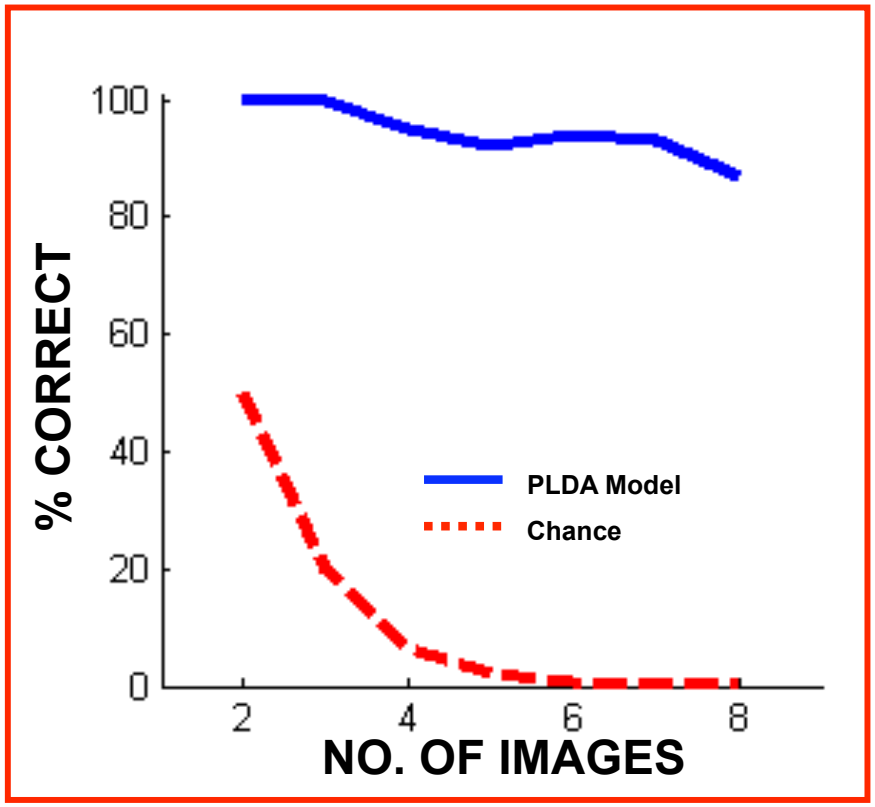
Extract normalized vector of pooled gradients around each feature

Build a PLDA model of each

# Model Explosion

| IMAGES: | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| MODELS: | 2 | 5 | 15 | 52 | 203 | 877 | 4140 |

For small N, we can compare all hypotheses



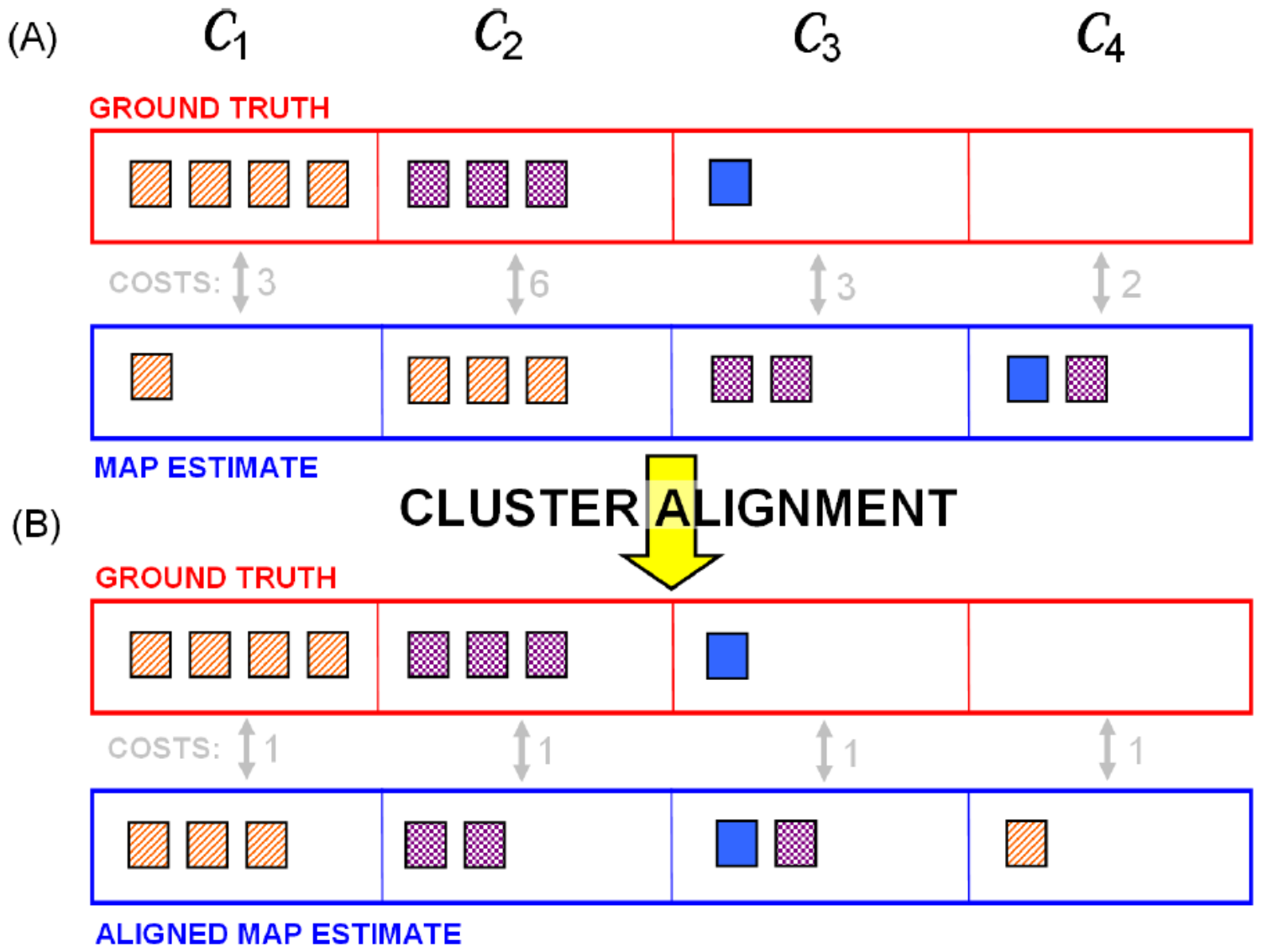

**PROBLEM 1:** For larger N, the number of models is too large ($10^{115}$ for 100 images) and we must resort to approximate methods.

Agglomerative approximation used – start with N different groups and merge the best two repeatedly until the likelihood stops increasing

**PROBLEM 2:**

We almost never get the answer exactly right when N=100. % Correct is not a good metric. How should we measure clustering performance

# Cluster Alignment



# Clustering Metrics

- Rand Metric

$$E_1(\mathcal{M}, \mathcal{M}') = \frac{N_{11} + N_{00}}{n(n-1)/2}$$

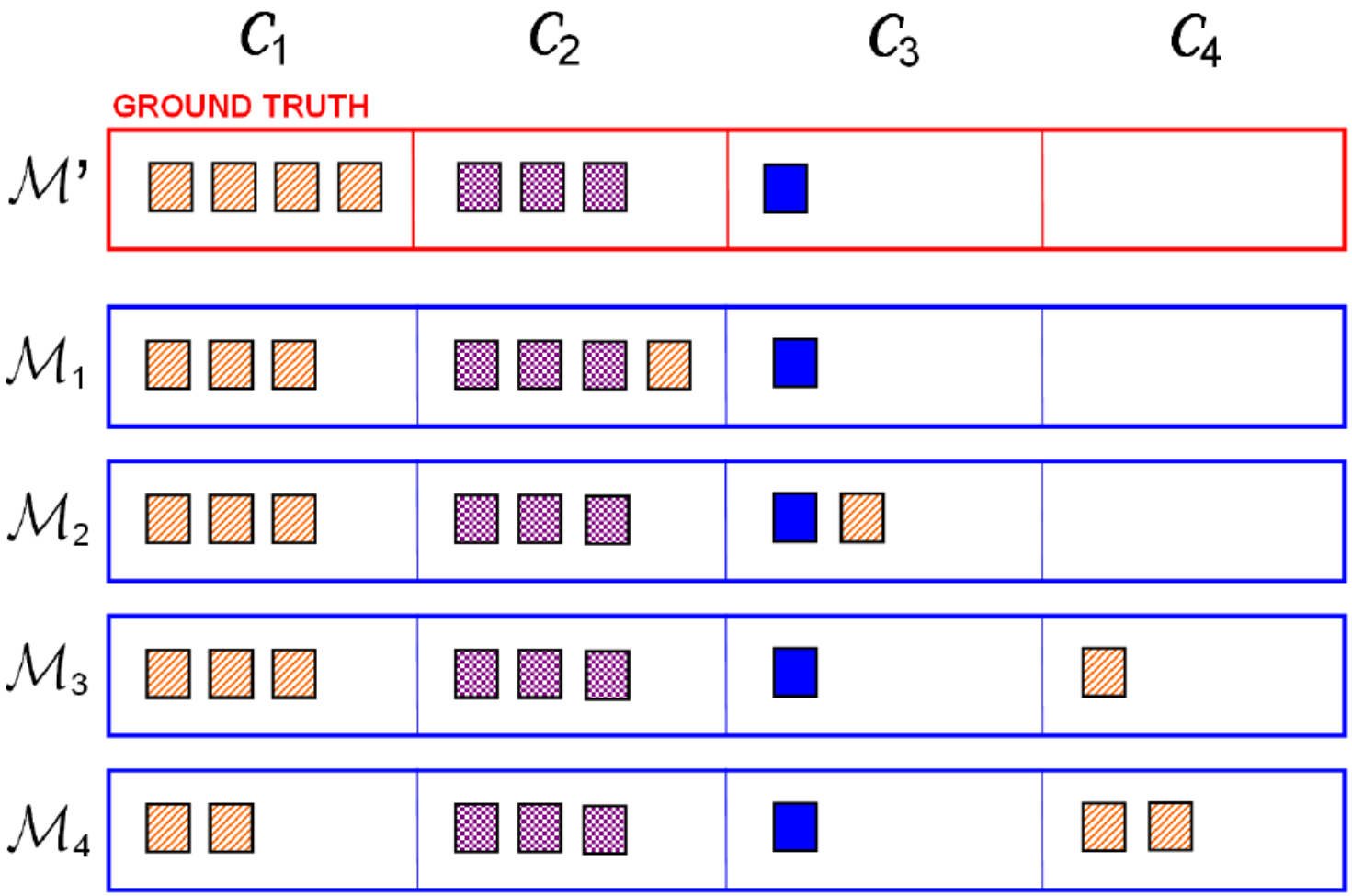
- Variation of Information

$$E_2(\mathcal{M}, \mathcal{M}') = H(\mathcal{M}) + H(\mathcal{M}') - 2I(\mathcal{M}, \mathcal{M}')$$

- Precision and Recall

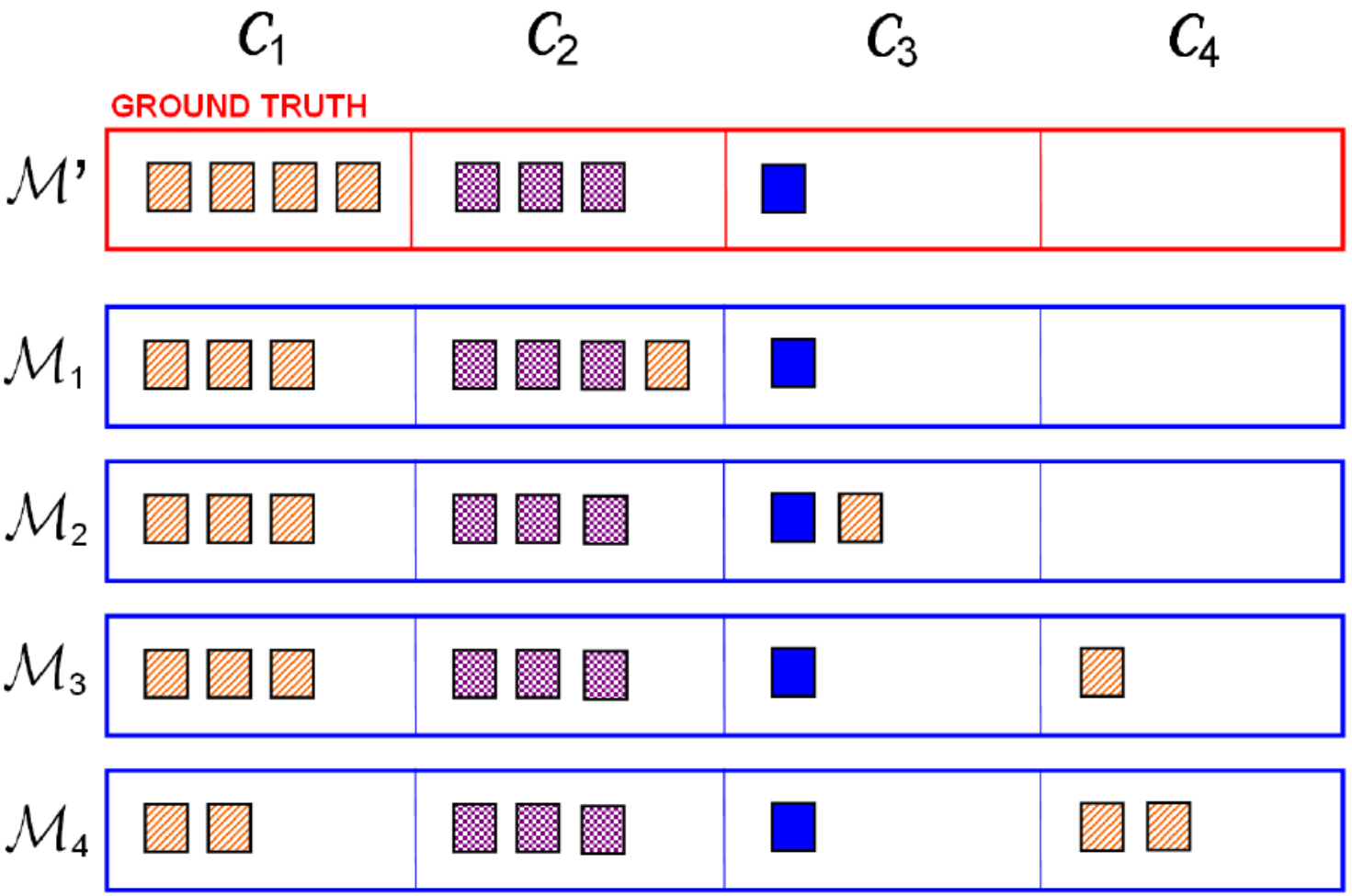
$$E_3 = \frac{\sum_{k=1}^{N_c} n_k \text{Precision}(k) + n_k \text{Recall}(k)}{2\sum_{k=1}^{n_c} n_k}$$

# Which clusters are the best?




Surely $M_1=M_2<M_3=M_4$?

# Which clusters are the best?




Propose using no of splits and merges required to match

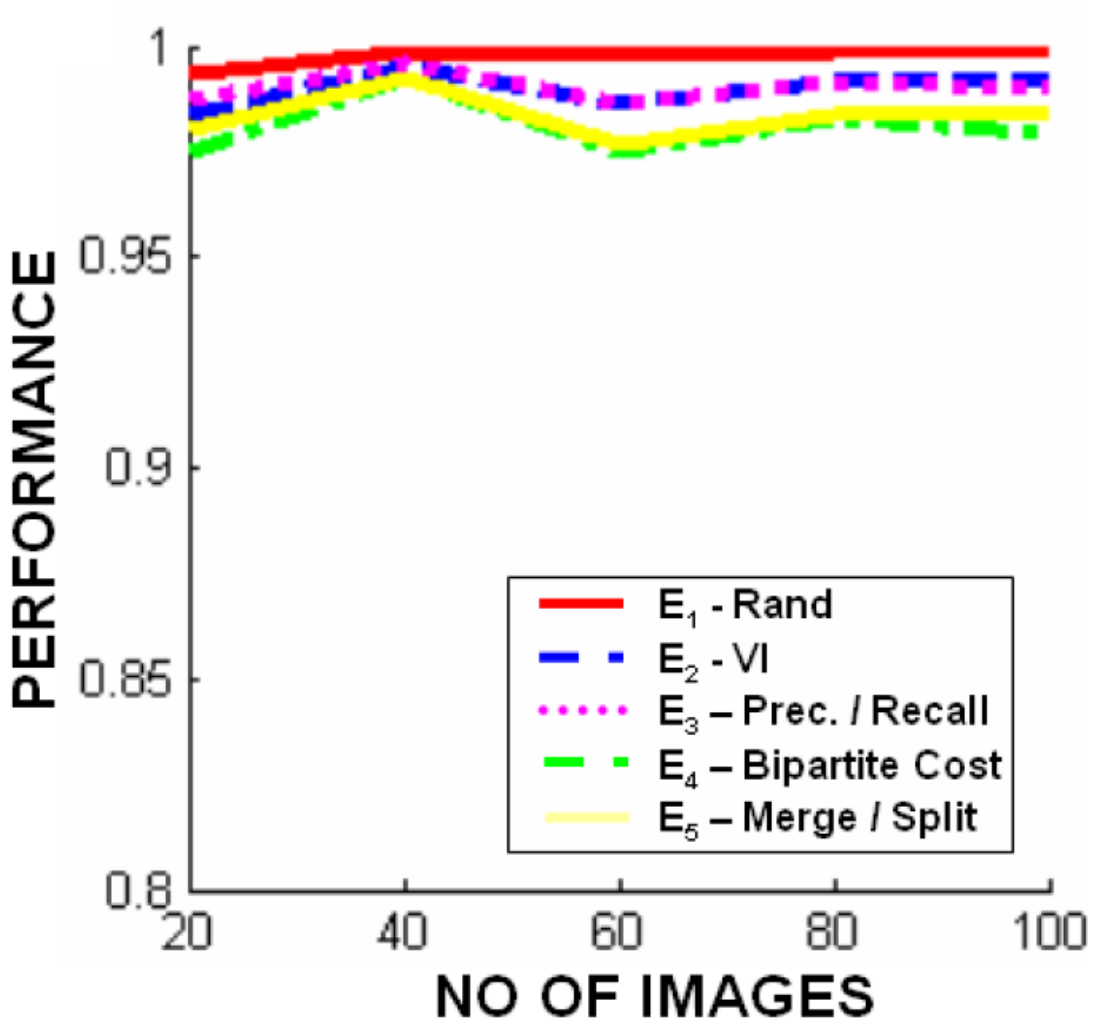
# Quantitative Results

XM2VTS Frontal Data

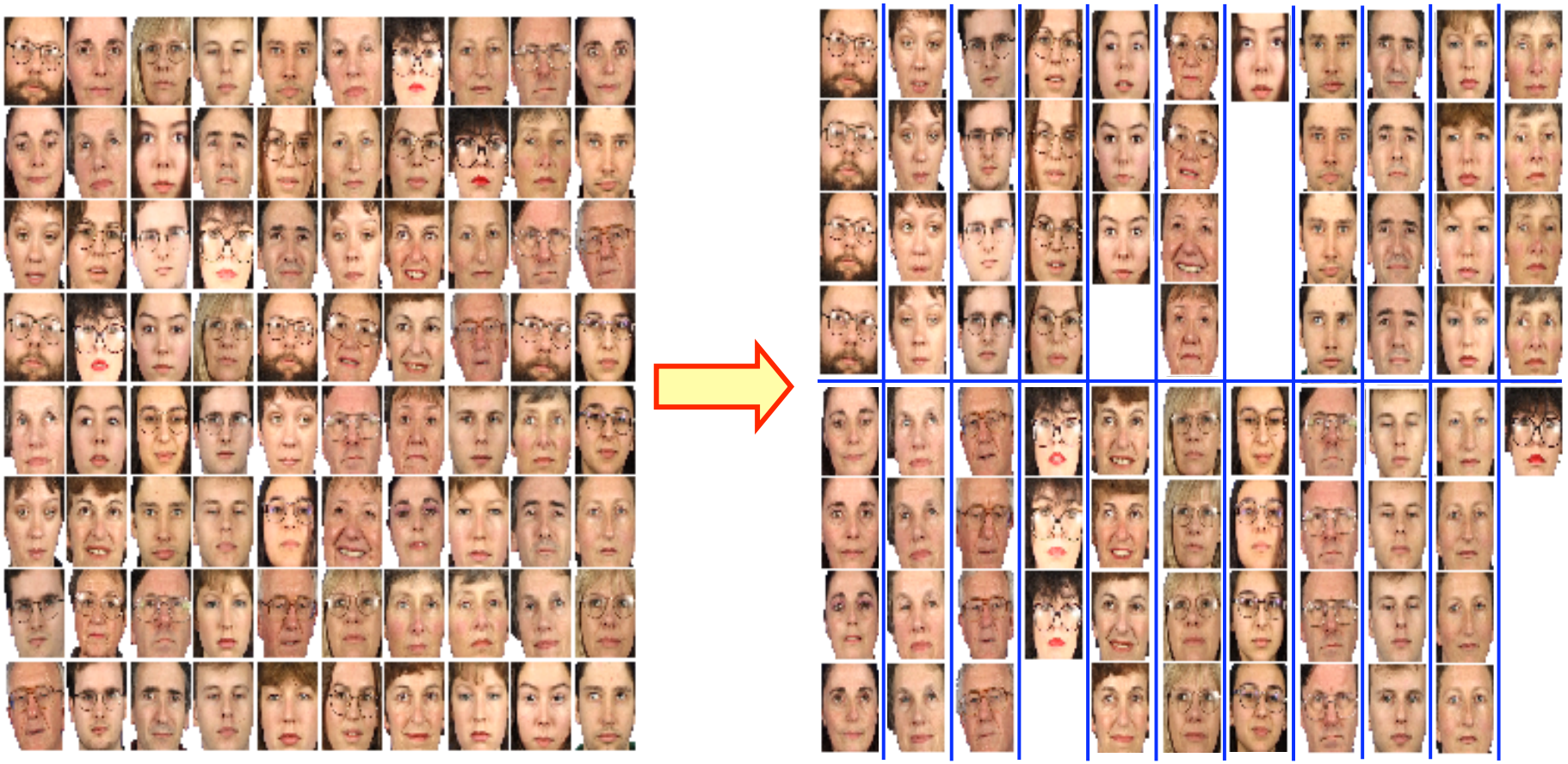
Always 4 images per person.

5,10,15,20,25 people

100 repeats for each datapoint with different people and different images



# Qualitative Results



Split-Merge cost here = 0.97  (average was 0.98)

# Cross pose case

What if there is more than one pose?

# Tied PLDA

$$\mathbf{x}_{ijk} = \boldsymbol{\mu}_k + \mathbf{F}_k \mathbf{h}_i + \mathbf{G}_k \mathbf{w}_{ijk} + \epsilon_{ijk}$$

Observed data from j th image of i th Individual in k th pose

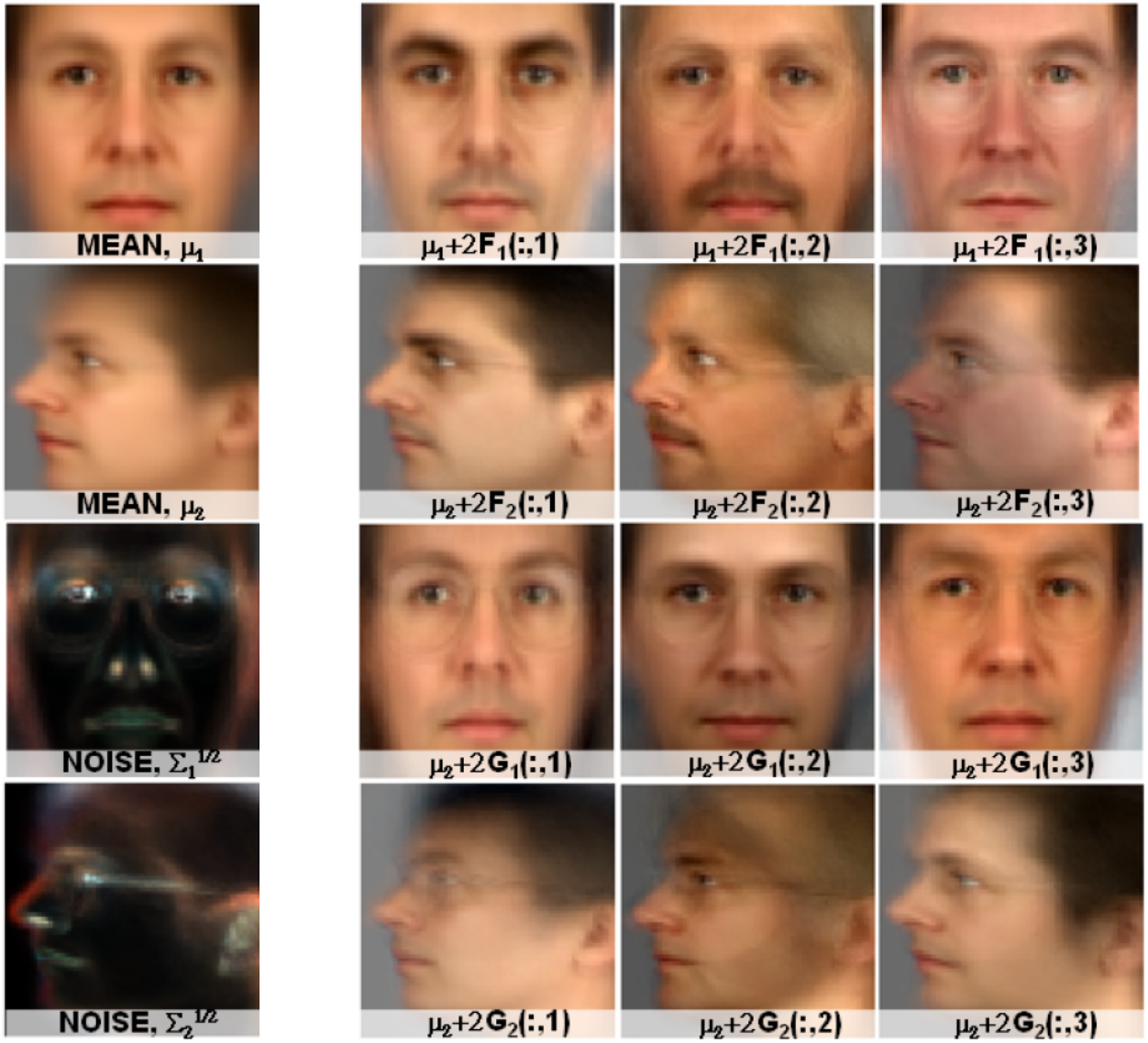
Overall mean for pose k

Weighted sum of basis functions **F** for between individual variation (identity) for this pose

Weighted sum of basis functions **G** for within-individual variation for this pose

noise

# Tied PLDA




Between-individual variation (pose 1)

Between-individual variation (pose 2)

Within-individual variation (pose 1)

Within-individual variation (pose 2)

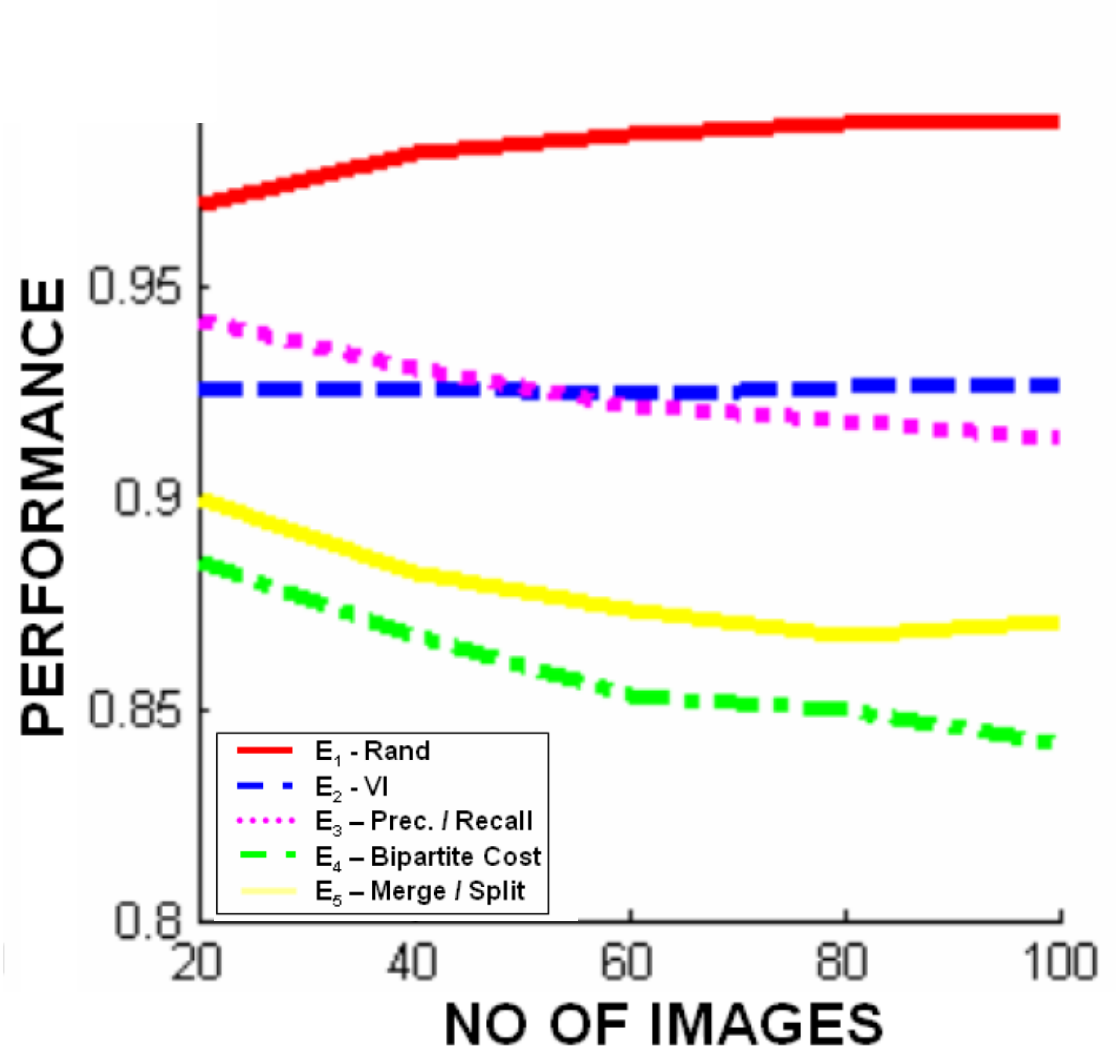
# Quantitative Results

XM2VTS Frontal Data

Always 4 images per person.

5,10,15,20,25 people

100 repeats for each datapoint with different people and different images



# Qualitative Results

Split merge cost = 0.85 (average 0.865)

# Conclusions

- Clustering of faces is a difficult problem
  - Multiple face recognition
  - Choice of model size
- Our method marginalizes over number of identities – can compare models of different size
- Extends to the cross-pose case

Learn more about these techniques via http://pvl.cs.ucl.ac.uk