

1



Statistical Facial Identification

Nick Fieller

joint work with Lucy Morecroft, Ian Dryden
& *al* of the IDENT project (PI: Martin Evison)

Department of Probability & Statistics
University of Sheffield, UK

4th October 2006

Statistical Facial Identification, Unliver, 04/10/06

2

- **Project** Funded by FBI via US Government
 - ◆ 2003 – 2005





Richard Vorderbrugge



- Joint project with Universities of Nottingham & Kent
- Principal Investigator Martin Evison
 - Former Senior Lecturer in Forensic Pathology in Sheffield




Statistical Facial Identification, Unliver, 04/10/06

3



- **Objectives of Facial Identification**
 - ◆ To provide **quantifiable** measures of quality of match of faces
 - ◆ Use as **evidence** in a Court of Law
 - e.g. Face captured on CCTV camera
 - Suspect is arrested and photographed
 - What is **probability** of the two pictures of faces being of the **same** person?
- **NOTE:**
 - ◆ This is different from **facial recognition**
 - *What is the best match in a database to a face in the crowd*

Statistical Facial Identification, Unliver, 04/10/06

4

- **Facial Recognition**
 - ◆ People are very good at recognising faces
 - » (perhaps too good)
 - ◆ Common experience of many people that they 'recognise' somebody at a distance but later realise it is somebody else
 - ◆ Has anybody ever mistaken anybody in the street for *their own mother*?
 - ◆ Case of Delimar Vera Cuevas:–
 - Mother recognised child she had not seen for 6 years

Statistical Facial Identification, Unliver, 04/10/06

5

■ Delimar Vera Cuevas

Mother finds kidnapped child six years on

A 10-day-old girl thought to have died in a fire in 1997 was actually kidnapped by a woman who started the blaze to cover her tracks, police said yesterday.

The girl's real mother saw the girl, now six, at a birthday party and recognised her as her own.

Delimar Vera was thought to have died in her family's home in Philadelphia, consumed by the heat and flames of a fire blamed on an extension cord for a heater. No body was ever found.

Captain John Darby of the Philadelphia police said the mother, Luz Cuevas, contacted the authorities after spotting the child in January. An investigation prompted DNA tests that confirmed her sus-



Delimar Vera: recognised by mother at birthday party

picion. The mother "didn't know whether to cry, to yell or to scream", a police officer, Manuel Gonzalez, said. "She was just in total shock."

Police have issued a warrant for the arrest of Carolyn Correa, 41, of Willingboro, New Jersey, on charges of arson, kidnapping and conspiracy. Her whereabouts were unknown.

"This child, now six years old, who has been raised by Carolyn Correa as her own, is not her own," Capt Darby said.

Ms Cuevas told a local television station that she recognised the child from a glimpse on her face.

"I said to my sister, look, she's my daughter," Ms Cuevas said.

It was unclear what brought

Statistical Facial Identification, Unliver, 04/10/06

the child and her mother to the same party.

State representative Angel Cruz, who helped Ms Cuevas to contact the police after she spotted the girl, credited "motherly instinct" for connecting the parent and child.



Ever since the blaze, Ms Cuevas had held on to the belief that her child was alive, partly because it did not make sense that a window in the infant's room was found to have been open even though it was the middle of December, Mr Cruz said.

The girl was placed in the custody of New Jersey division of youth and family services. It was not clear when she would be reunited with her mother. AP, Philadelphia

Gustin Wad 3/3/4

6

- **The case of Delimar Vera:–**
 - ◆ 1997:– 10-day old baby Delimar believed burnt in house fire
 - ◆ 2004:– Mother recognises child at a birthday party in January
 - ◆ 2004:– February; DNA tests on sample of hair taken by mother prove identity
 - ◆ 2004:– March; Carolyn Correa, who had raised Delimar for six years, charged with arson + kidnapping + conspiracy
- **Unanswered question:–**
 - ◆ **Both at same party only a coincidence?**





Statistical Facial Identification, Unliver, 04/10/06



7

A mother's instinct



December 1997 January 2004



Statistical Facial Identification, Unlver, 04/10/06



8




Statistical Facial Identification, Unlver, 04/10/06



9



A mother's instinct is impressive, — but see later



Statistical Facial Identification, Unlver, 04/10/06



10


<http://www.youlooklike.com>



11

FRANÇOIS BRUNELLE


BIOGRAPHY
I'M NOT A LOOK-ALIKE!
PRESS
FAQ
CONTACT




SARAH FOURNIER
AND ALAN MADILL
TORONTO, CANADA, 2001
+ >>

I'm not a look-alike! is a collection of François Brunelle's photographic portraits of North American and European look-alikes. Each photo features two look-alikes side by side, their posture neutral, set against a uniform background.

TO PARTICIPATE AS A LOOK-ALIKE



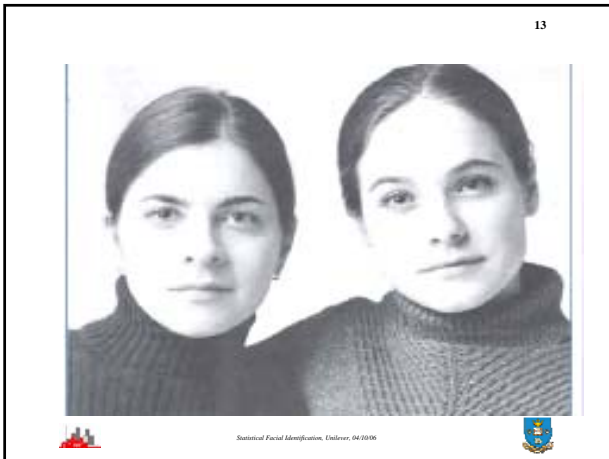
Statistical Facial Identification, Unlver, 04/10/06



12





14

- Facial **recognition** by pattern recognition
 - ◆ algorithms for matching shapes, shades & contours of a face of known individual with photographic matches from CCTV

Depending on the tolerance set, may match to wrong face or even a non-face on image, e.g. dog matching lighting patterns

Statistical Facial Identification, Unliver, 04/10/06

15

- Generally, automatic computer facial recognition works well
 - ◆ Screen crowds in football matches
 - ◆ Monitor airports
 - ◆
- Recognised faces are then questioned
 - ◆ Facial recognition not used as evidence only used to pick out a few key suspects
 - ◆ Match of photographs not used as main legal evidence
 - same with Delimar Vera where proof came from DNA not photographs
 - c.f. fingerprints

Statistical Facial Identification, Unliver, 04/10/06

16

- People believe that faces are unique
 - ◆ Identification by a witness is often impressive evidence in a court of law
 - ◆ Either pick suspect from line of 12 or from 8 DVDs or personal declaration
 - ("p-value" ~ 0.08?? or 0.12???)
 - ◆ People feel they 'understand' faces
 - even used in statistics as Chernoff Faces
- But people do not **'identify'** faces — they **'recognise'** faces

Statistical Facial Identification, Unliver, 04/10/06

17

- **Recognition:-**
 - ◆ Making a **match** between a new image and a set of images in a database (or in memory)
 - ◆ If database (or memory) does not have a true copy of new image then 'best available' is selected
 - Leads to cases of *mistaken identity*
 - ◆ e.g. case of Peter Hain, (& others recently).....

Statistical Facial Identification, Unliver, 04/10/06

18


Lines drawn on photographs joining features: if lines on different photographs joining same features intersect common features in both photographs then evidence of a 'match'

Statistical Facial Identification, Unliver, 04/10/06




19

- **Outline of Our Approach**
 - ◆ Identify 'landmarks' on photos of faces
 - coordinates of identifiable features
 - corner of eyes, tip of nose,
 - ◆ Extract 'shape' of configuration of landmarks
 - 'shape':– property not affected by uniform changes in size, orientation/rotation,...
 - statistical analyses can do this
 - Procrustes analysis
 - Ian Dryden




Statistical Facial Identification, Unliver, 04/10/06




20

- **Shape analysis**
 - Key reference: Dryden, I. L. and Mardia, K. V. (1998) 'Statistical Shape Analysis', Wiley.
 - Simplest is 2D:
 - Take coordinates $\{(x_j, y_j) ; j=1, \dots, k\}$ of landmarks as points in complex space $z_j = x_j + i y_j$ [$i = \sqrt{-1}$]
 - Treat values of z_j as values of k complex variables
 - Calculate variance matrix of the z values for k objects
 - First eigenvector (i.e. principal component) gives mean Procrustes corrected shape
 - Can then examine variability of shapes along directions of principal components
 - Public domain software (from R library) for Procrustes Analysis in 2D & 3D & fitting of models for shape distributions





Statistical Facial Identification, Unliver, 04/10/06




21

- **Assess variations in measured shapes**
 - ◆ Sources of variation:
 - 1: variability **between** different faces (i.e. people)
 - 2: variability **within** people (ageing, expression,...)
 - 3: variability **between** different operators measuring shapes from same photo
 - 4: variability **within** operator making different measurements of same photo
 - 5: different photos of same face,
 - ◆ Initial 2002 study shows source 1 typically much greater than sources 3 & 4





Statistical Facial Identification, Unliver, 04/10/06




22

- **Provides a route for matching people**
 - ◆ if 'shapes' are closer than population variability would suggest then indicates a match
 - ◆ variability can be quantified
 - ⇒ probability of match
 - ◆ If we have statistical models for shape distributions (see references) then can update evidence via Bayes factors or Likelihood Ratios
 - See Aitken & Lucy (2004), *Applied Statistics, (JRSS, series D)*

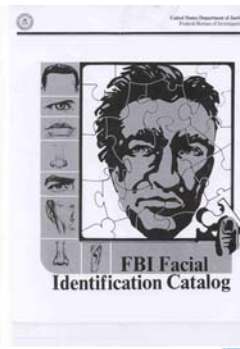



Statistical Facial Identification, Unliver, 04/10/06




23

- **Pilot Study**
 - ◆ Based on measuring faces from the FBI facial identification catalogue
 - ◆ Faces scanned & landmarks captured manually





Statistical Facial Identification, Unliver, 04/10/06





24

AVERAGE EYES




CLOSESET EYES





Statistical Facial Identification, Unliver, 04/10/06




25

(it was not Al Capone)

Which face has appeared twice?

Statistical Facial Identification, Unlver, 04/10/06

26

Traditional anthropometric landmarks

coordinates captured 'manually' with public domain software

Statistical Facial Identification, Unlver, 04/10/06

27

Four subjects from the 2002 pilot study

Different sections of the face are censored so only use landmarks common to a pair of faces

Statistical Facial Identification, Unlver, 04/10/06

28

Cluster analysis of Procrustes coordinates of mean shapes of 48 images from 2 sections of catalogue: matches identified

Statistical Facial Identification, Unlver, 04/10/06

29

Statistical Facial Identification, Unlver, 04/10/06

30

- Main Study
 - ◆ 3D Scanning of ~ 3,000 faces
 - ◆ Assess population variability
 - ◆ Determine landmarks from both 3D scan and 2D full-face picture
 - ◆ Match oblique photo to 3D scan
 - Real-life situation:
 - Oblique photo from surveillance camera
 - 3D scan from suspect

Statistical Facial Identification, Unlver, 04/10/06



31

■ **3D scan taken from eight 2D digital photos**




Lucy with Geomatrix® 3D Scanner





Statistical Facial Identification, Ulster, 04/10/06

32

■ **The eight 2D views**





Generates 3D pictures in any orientation to match view from crime scene





Statistical Facial Identification, Ulster, 04/10/06

33

Cyberware laser scanner for verification of Geomatrix model




Statistical Facial Identification, Ulster, 04/10/06

34

■ **Preliminary Study**

- ◆ **Five** assistants of 3D scanner
- ◆ **Two** operators taking **61** landmark coordinates
 - ⇒ inter/intra observer reproducibility study
 - 35 people scanned by one of the five assistants
 - each scan landmarked by each operator three times
 - ⇒ allowed choice of 'best' landmarks for main study
 - 'best' ≡ 'reproducibility/discriminatory/reliability'
 - consistency of placement/conventional authority
- ◆ **Validation** of 3D scanner with 3D laser and direct calliper measurements
 - some concerns on accuracy of Geomatrix® software




Statistical Facial Identification, Ulster, 04/10/06

35

■ **Main Study**

- ◆ ~ 3000 3D scans taken of public volunteers
 - all landmarked twice
 - ('best' 30 landmarks)
 - using Geomatrix® scans
 - » Cyberware laser scans taken of most subjects but not used
- ◆ ~ half of subjects provided complete sets of 30 landmarks
 - e.g. hair obscuring ears, facial hair.....
 - ~8000 values 'missing' out of 90000 (=30×3000)
 - (or 24000 from 270000 since 3D coordinates)
 - Estimate by EM algorithm (iteratively *ad hoc*)
 - assuming [singular] MVN model for [Procrustes corrected] 3D coordinates




Statistical Facial Identification, Ulster, 04/10/06

36

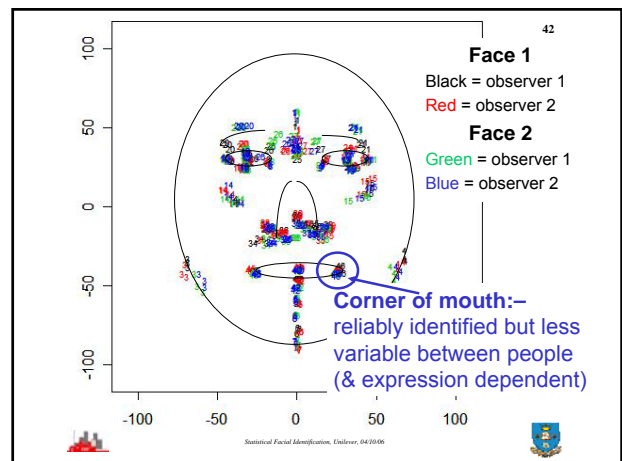
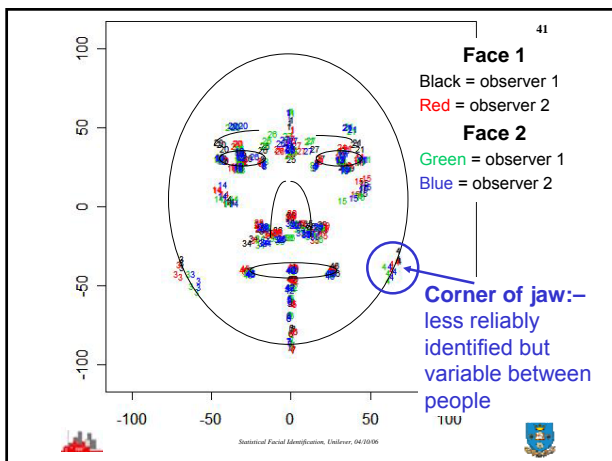
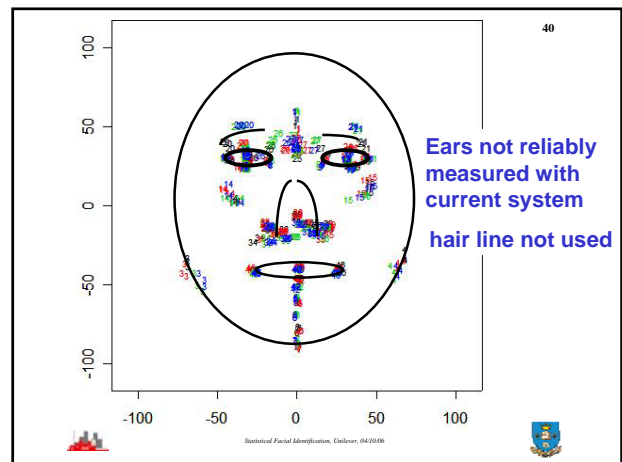
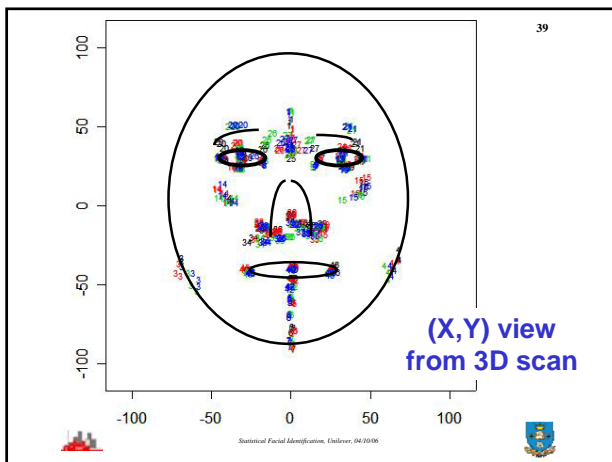
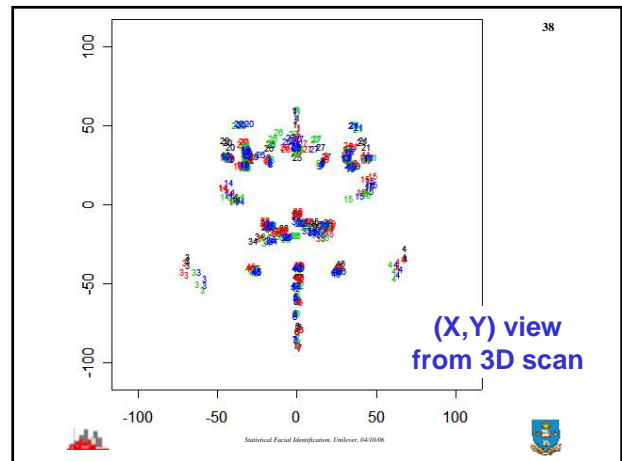
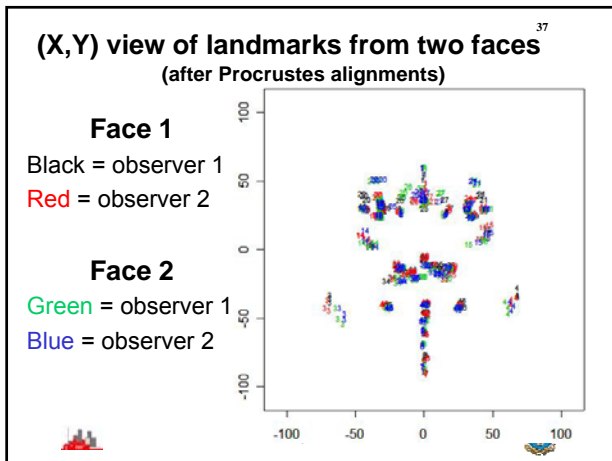
■ **Reliability of landmarks**

- ◆ need landmarks which are:-
 - reliably identifiable by different observers
 - variable between different people
- **Illustration:-**
 - ◆ 2 faces each measured by 2 observers
 - 3 replicates of each landmark




Statistical Facial Identification, Ulster, 04/10/06






43

- **Full study**
 - ◆ 35 different faces from database
 - ◆ 61 landmarks
 - ◆ 2 operators, each placed landmarks 3 times
 - (random orders of subjects)
 - ◆ Procrustes Alignment carried out
 - Looked at:-
 - Consistency of each observer
 - Observer judgement of landmark location
 - (did it vary between observers?)
 - ◆ Plots show scores on first few PCs of the residual values after subtracting an overall mean from the Procrustes aligned configurations

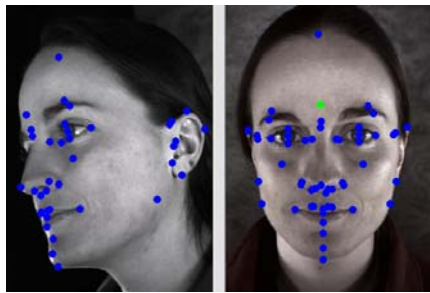



Statistical Facial Identification, Unliver, 04/10/06




44

61 original landmarks

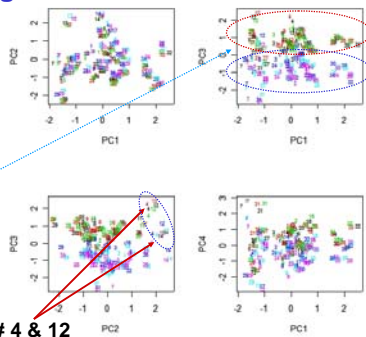
Statistical Facial Identification, Unliver, 04/10/06




45

PCA on Procrustes aligned data using 61 landmarks


- Observer 1
 - ◆ Black, red, green
- Observer 2
 - ◆ Blue, cyan, pink
- Numbers represent faces 1-35
- Systematic difference between observers on PCs 3 and 4
- Much more than difference between individuals



Faces # 4 & 12




Statistical Facial Identification, Unliver, 04/10/06




46

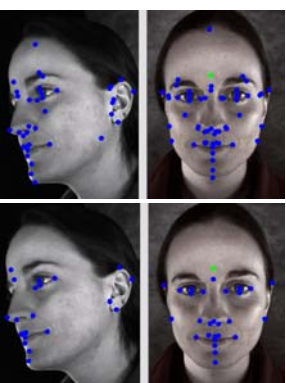
- Landmark list reduced
 - ◆ Factors considered:
 - observer knowledge
 - consistency of landmark placement
 - discriminatory power
 - landmark visible in majority of cases?
- Final list with **32 landmarks**
 - (later down to 30)
 - ◆ cf **reported** plans for Biometric Passports to use only 16 landmarks
 - Is this correct????? Anybody know???



Statistical Facial Identification, Unliver, 04/10/06




47




Original 61 landmarks

Selected 32 landmarks



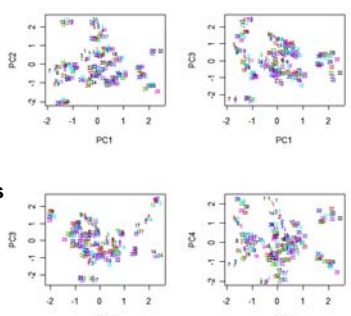

Statistical Facial Identification, Unliver, 04/10/06




48

PCA on Procrustes aligned data using selected 32 landmarks

- Observer 1
 - ◆ Black, red, green
- Observer 2
 - ◆ Blue, cyan, pink
- Numbers represent faces 1-35
- Little systematic difference between observers on 1st 4 PCs
- Suggests these landmarks good for reproducibility and discrimination





Statistical Facial Identification, Unliver, 04/10/06



49

■ **Back to Delimar Vera**



Delimar as a baby:
15 anthropometrical landmarks (2D) have been located on this image (2 repetitions of each).

Statistical Facial Identification, Unlver, 04/10/06

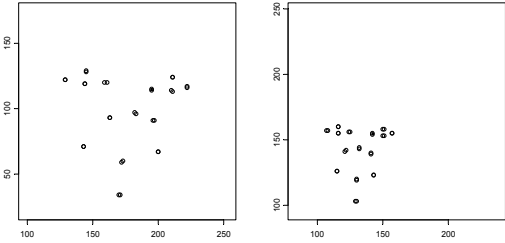
50



Delimar at six years old:
The same fifteen landmarks (2D) have been located on this image (2 repetitions of each).

Statistical Facial Identification, Unlver, 04/10/06

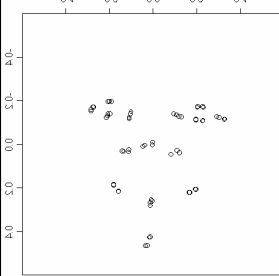
51



Plots of the raw landmark coordinates (left is the baby picture, right is the picture from six years old). Quite clearly the scales for the two images are not the same.

Statistical Facial Identification, Unlver, 04/10/06

52



Plots of the Procrustes rotated coordinates from both images (baby & six years old)

.....a fairly close match?

Assessment based on modelling coordinates as MV normal, estimated covariance from sample data. Clearly important to include gender/age group etc if possible.

Statistical Facial Identification, Unlver, 04/10/06

53

■ **Various statistical issues:-**

- ◆ **Missing values**
 - ~ 10% missing, imbalanced pattern, possibly informatively missing
 - ad hoc imputation by EM algorithm piecemeal based on various complete subsets
 - certainly worthwhile iteratively updating
- ◆ **Which landmarks are best (& how many) ?**
 - Conflicting requirements of discriminatory power vs reliability of measurement
 - selection methods used so far are pragmatic
 - » (c.f. stepwise selection methods in other contexts)

Statistical Facial Identification, Unlver, 04/10/06

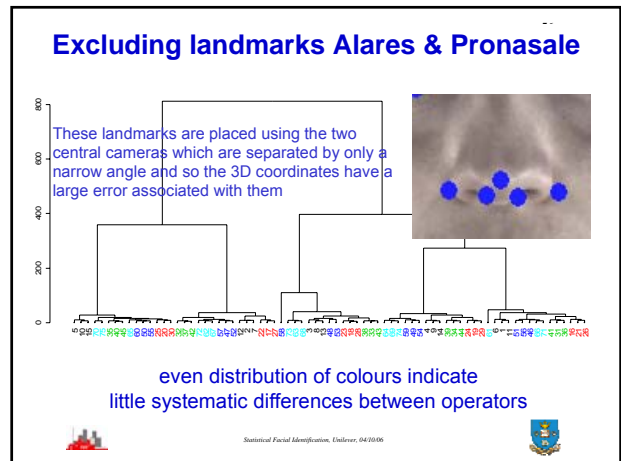
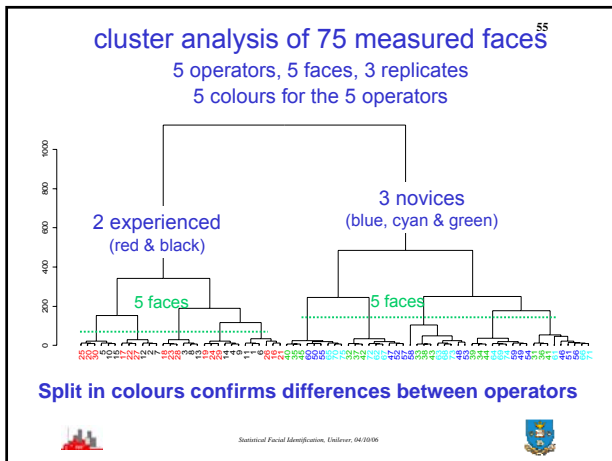
54

■ **How well will new operators perform?**

- ◆ **are some landmarks troublesome for novices?**
- ◆ **2 experienced and 3 new operators**
 - each measure 5 faces 3 times
 - $5 \times 5 \times 3 = 75$ sets of measures
 - cluster analysis of 75 sets
 - (single link, Euclidean distance of Procrustes coordinates)
 - close agreement between two experienced operators
 - close agreement between within-operator replicates of same face
 - clear separation between operators
 - clear separation between different faces

Statistical Facial Identification, Unlver, 04/10/06



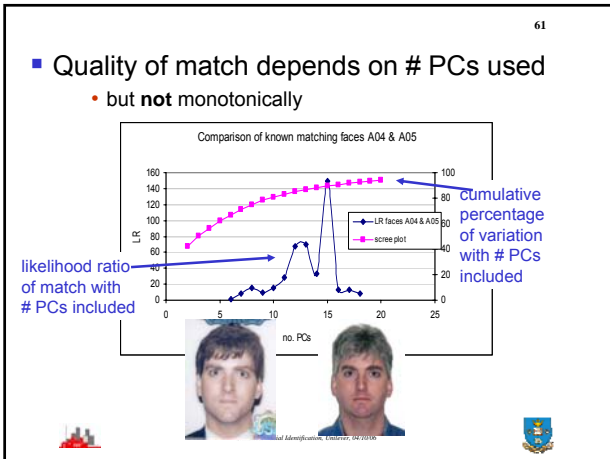


- 57
- How much evidence do we need to exclude or admit a match?
 - ◆ Working models based on assumption of multivariate normal observations $X_i \sim N_p(\mu, \Sigma)$
 - estimate μ and Σ from the data in the usual way
 - based on sample values from collected 3000 faces
 - » may be worth correcting for sex, age, ethnicity,...
 - If X_i are raw landmarks then $p \sim 90$
 - 30 landmarks, 3 dimensions (as in EM imputation)
 - smaller if corrected for size and rotations
 - » (i.e. Procrustes residuals)
 - smaller if data referred to first k principal components
 - » say in range 5 - 30?
 - smaller if working with 2D images
 - MV normal seems reasonable
 - better with first few PCs

- 58
- Assess evidence by 'Likelihood Ratio'
 - ◆ X_1 data from person P_1 from crime scene
 - X_2 data from person P_2 who is suspect
 - ◆ interest in $H_0: P_1 = P_2$ vs $H_1: P_1 \neq P_2$
 - not appropriate to use formal statistical test
 - » (e.g. Hotelling's T^2)
 - ◆ Assess by $LR = f(x_1, x_2 | P_1=P_2) / f(x_1, x_2 | P_1 \neq P_2)$
 - see Aitken & Lucy (2004, Applied Statistics)
 - high LR supports evidence of guilt of suspect
 - low LR decreases evidence
 - could incorporate with priors & other evidence as a Bayes factor
 - ◆ How high do need the LR to be impressed?
 - ('beyond reasonable doubt')

- 59
- Investigations with 2D images of Richard Vordebrugge and some known & test cases
 - ◆ Certain people produced 'matches' with many different people (with $LR > 1$)
 - ◆ maybe some people have very 'average' faces (c.f. Gary Lineker)
-
- these 'matched' with each other & several others

- 60
- Quality of match depends on # PCs used
 - but **not** monotonically
-



- 62
- Many other problems
 - imbalanced MV data
 - variable selection
 - what facial features are best for identification
 - (immune to facial expression but distinctive between people?)
 - strength of evidence
 - how much to presume re covariates
 - e.g. ethnicity, sex,.....
 - ...
 - ...
- <http://nickfieller.staff.shef.ac.uk>
- Statistical Facial Identification, Unlver, 04/10/06

