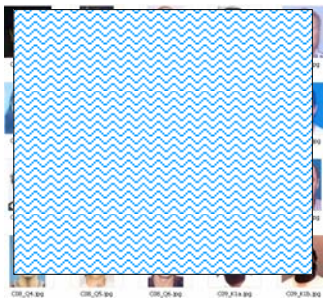
Case Study:– Interpretation of PCs

- slides & work courtesy of **Lucy Morecroft**

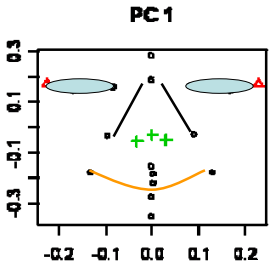
- **Data:-** Coordinates of facial landmarks
  - ~ 90 variables, ~3000 faces
  - Initial analysis:-
    - Standardize ***size & orientation*** of faces
      - » (so no component attributable to variation in sizes)
  - Preliminary analysis:
    - Refer data to first **k** principal components
      - » (k < 90, k ≈ 30????)
    - Fit a multivariate normal model $\mathbf{N_k(\mu, \Sigma)}$
- **Objective:-**
  - Measure match between 2 pictures



- LR Multiple photos of subjects taken in different locations
  - e.g. two driving licences from different locations
    - (*& different names*)
- Can we measure the likelihood that they are the same person?

- Method:- calculate Likelihood Ratio using MVN model
  - Ratio of likelihoods of assuming two faces are (a) same (b) different
  - If LR > 1 then evidence of similarity, LR >> 1 then good evidence

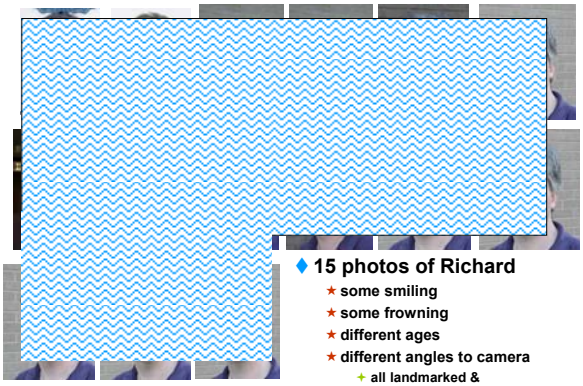
- Problems:-
  - How many PCs?
    - Presumably high order PCs only reflect noise and so should be discarded
      - (i.e. should not be used in MV Normal model)
  - Do we get better matches with more dimensions?
    - If we use 'too many' dimensions do we introduce noise and so reduce evidence????
    - What about ephemeral facial features???????
      - e.g. smiles, frowns — **temporary / short lived**
  - Need to know something about the PCs i.e. ***interpretations***



- Landmarks with dominant coefficients highlighted:
  - △ **< -0.2**
  - \+ > 0.2
  - O (-0.2 , + 0.2)
- **order of importance of features dependent on data base**
  - **(e.g. our subjects all kept straight faces so 'smile PC' is only at # 10)**

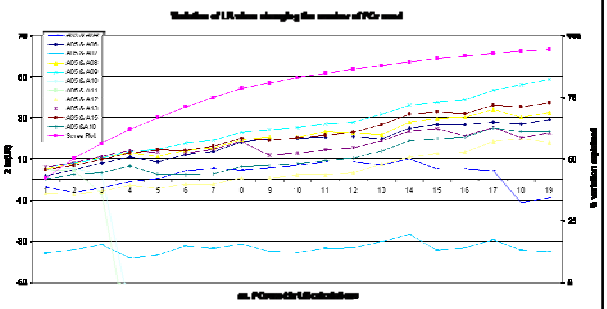
- allows interpretations of PCs:-
  - 1: outer eye breadth
  - 2: chin shape
  - ……
  - smile / …/ frown / …/wink /... / asymmetry / ……

- Test example:-
  - multiple photos of single subject
  - models fitted to successively increasing number of PCs
    - i.e. 1st to PC1, then (PC1, PC2), then (PC1, PC2, PC3)…
  - LRs of pairwise comparisons between images
  - Generally LRs increase with number of dimensions used in model, i.e. $LR_{k+1} > LR_k$
    - but some exceptions, i.e. $LR_{k+1} < LR_k$
  - **Which PCs *weaken* evidence of matches?**
- Plots show scree plot of % variation & corresponding values of LRs

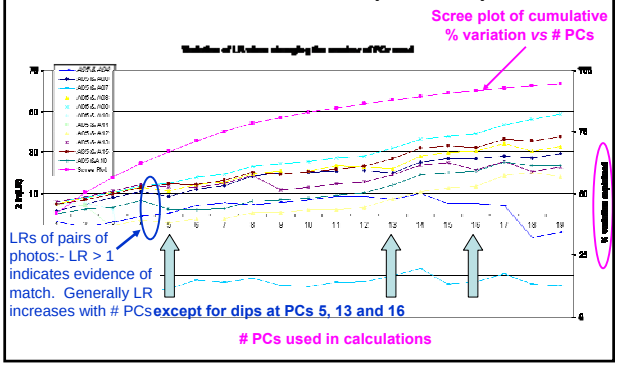


- **15 photos of Richard**
  - **some smiling**
  - **some frowning**
  - **different ages**
  - **different angles to camera**
    - **all landmarked & corrected for size**

How the number of PCs affects facial 'matches' (LR > 1)



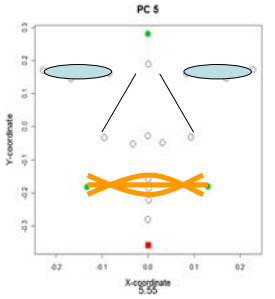
How the number of PCs affects facial 'matches' (LR > 1)



How does the number of PCs affects facial 'matches' (LR > 1)

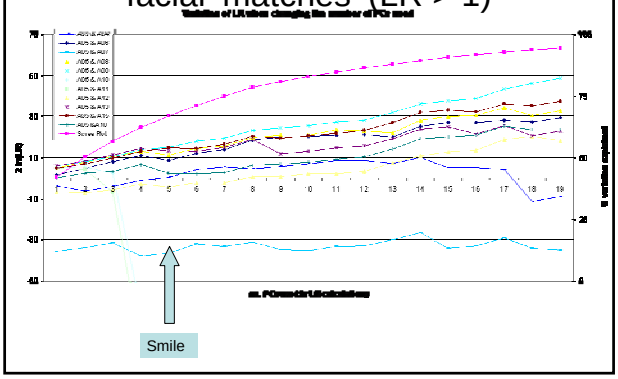
- as # PCs increases so does the LR & the evidence for a facial match (usually)
- Certain PCs show a decrease in strength of facial match – particularly PCs 5, 13, and 16
- PC loadings can be examined to see areas of the face that vary at each PC
  - i.e. their *interpretations*
    - Do they relate to features which should not be used in evidence?

PC 5 loadings plots

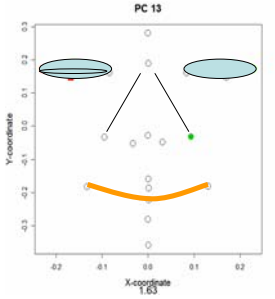


- PC 5 shows variation around the mouth
- This area may vary with facial expression
- If 2 images have different expressions the strength of evidence for a match decreases

How the number of PCs affects facial 'matches' (LR > 1)

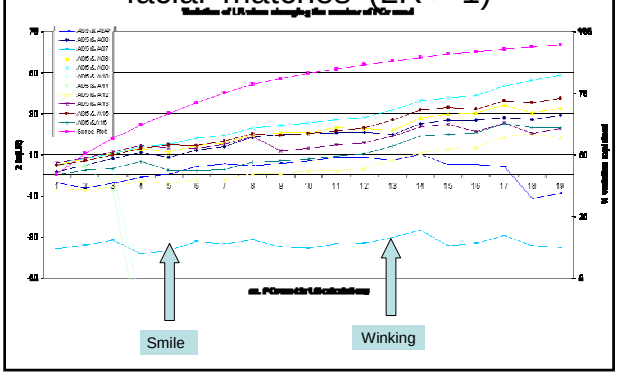


PC 13 loadings plots

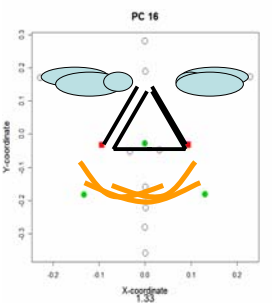


- PC 13 shows variation in shape of the left eye
- This area may vary if the subject winks or blinks
- Unless both subjects wink the strength of evidence for a match decreases

How the number of PCs affects facial 'matches' (LR > 1)

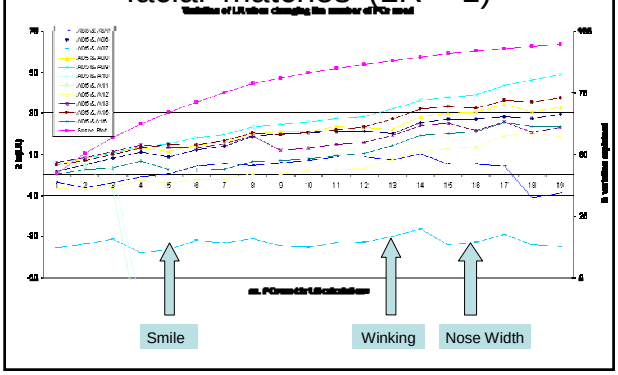


PC 16 loadings plots



- PC 16 shows variation in nose width
- This area may vary if the subject is not directly facing to the front
- Unless both subjects are positioned in the same 2D plane evidence for a match decreases

How the number of PCs affects facial 'matches' (LR > 1)



- Next steps & Summary:-
  - Try matching without the smiley, winking & nosy PCs
    - ★(ongoing work)
  - Initial analysis removed size variation
    - ★this would dominate PCA & is not 'interesting'
      - ✦(technique called 'Procrustes Analysis')
  - PCA partitioned remaining variation into components
    - ✦some components attributable to 'noise' & so were removed
    - ★identification of components allowed ephemeral features to be removed from assessment of evidence of matches