# 3D Facial Landmark Localization using Combinatorial Search and Shape Regression

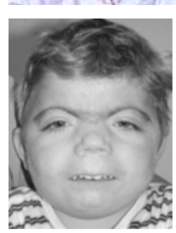
F.M. Sukno[1,2], J.L. Waddington[2] and Paul F. Whelan[1]
[1]Dublin City University and
[2]Royal College of Surgeons in Ireland

RCSI

DCU

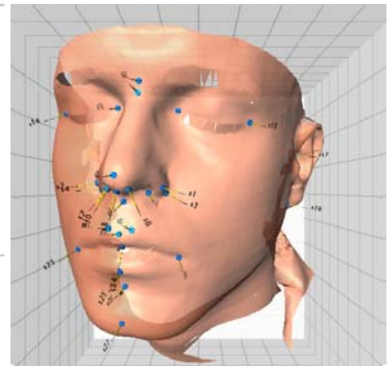
## Context

- Craniofacial geometry has been suggested as an index of early brain dysmorphogenesis in neuropsychiatric disorders
  - Down syndrome
  - Autism
  - Schizophrenia
  - Bipolar disorder
  - Fetal alcohol syndrome
  - Velocardiofacial syndrome
  - Cornelia de Large syndrome
  - Joubert syndrome
  - ...
- Patterns tend to be subtle



## Facial surface in 3D

- Larger availability of 3D imaging devices allows overcoming limitations inherent to 2D

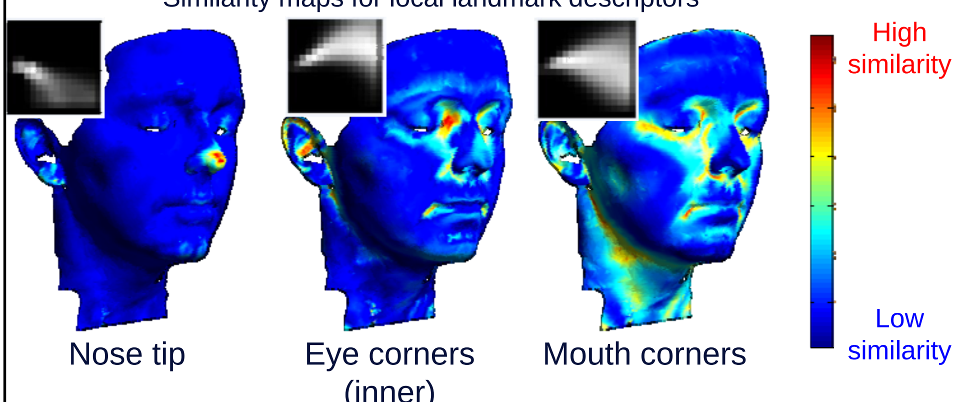


- Manual labelling of landmarks
  - Key points on facial features
  - Limited scalability, intra- and inter-observer variability

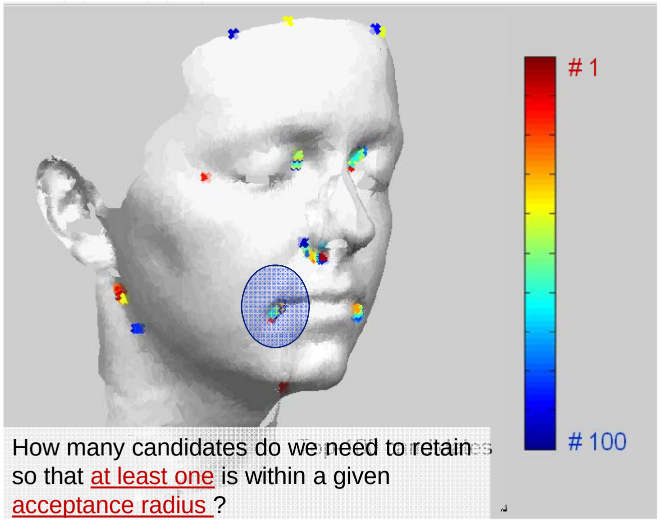
## Similarity maps with spin images

- Cross correlation of a template with every mesh vertex
- We start by identifying the top-candidates

Similarity maps for local landmark descriptors

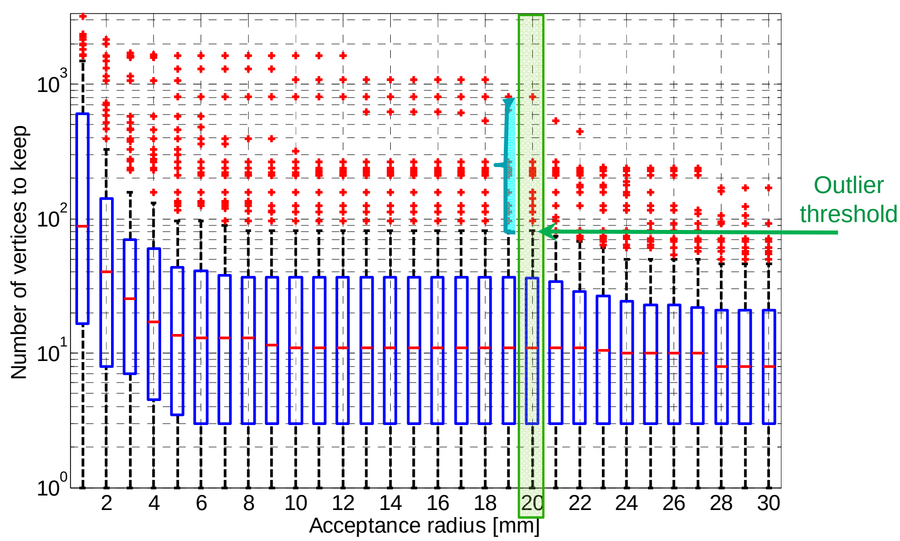


## Keeping the top-scoring vertices (*candidates*)



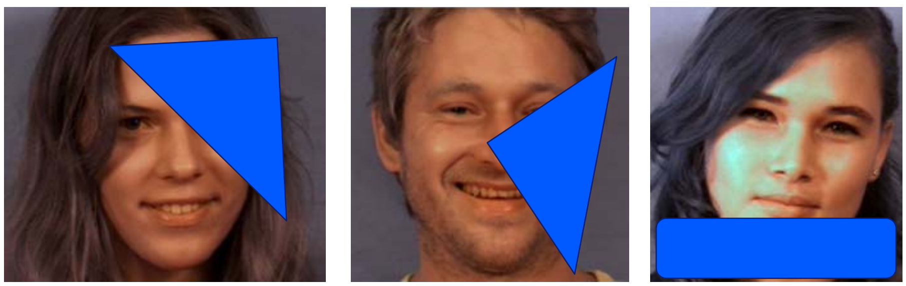

Example targeting the right mouth corner

## Dataset statistics: Example for the mouth corners



## Our approach

- Accept we will not find all landmarks (within retained candidates)
- Use statistical inference to complete missing landmarks
  - This allows reducing the number of candidates to retain
  - More landmarks can be found



## Statistical priors

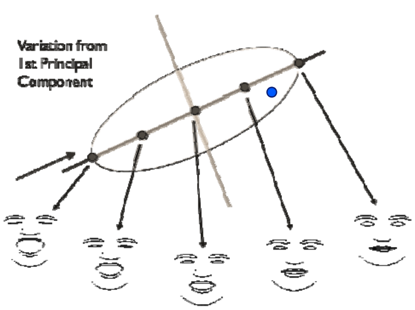
- Shape vector in 3D

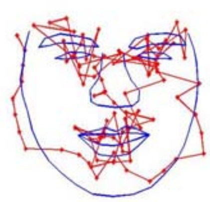
$$\mathbf{x} = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_L, y_L, z_L)^T$$

- PCA model from a training set

$$\mathbf{b} = \Phi^T(\mathbf{x} - \overline{\mathbf{x}})$$

$$\sum_{m=1}^{M} \left( \frac{b_m^2}{\lambda_m} \right) < \beta_e^2$$



## Shape regression with incomplete information

- We can group known or *fixed* coordinates and unknown ones (the ones the *guess*)

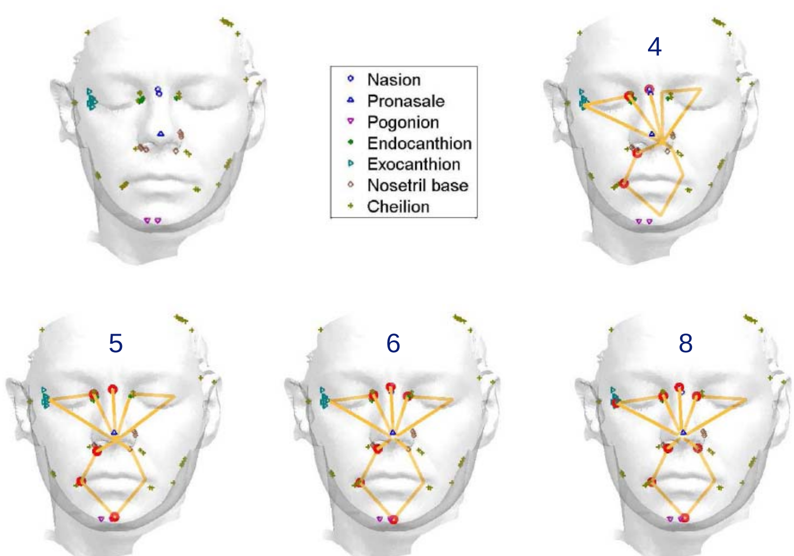
$$\mathbf{x}^g = (x_1, y_1, z_1, \ldots, x_g, y_g, z_g)^T \qquad \mathbf{x} = \begin{pmatrix} \mathbf{x}^g \\ \mathbf{x}^f \end{pmatrix}$$
$$\mathbf{x}^f = (x_{g+1}, y_{g+1}, z_{g+1}, \ldots, x_L, y_L, z_L)^T$$

- Assuming a multi-variate Gaussian distribution in shape space we find the coordinates that maximize the model probability:

$$Pr(\mathbf{x}) \sim e^{(-\mathbf{b}^T \boldsymbol{\Lambda}^{-1} \mathbf{b})} \qquad \frac{\partial Pr(\mathbf{x})}{\partial \mathbf{x}^g} = \mathbf{0} \Leftrightarrow \frac{\partial}{\partial \mathbf{x}^g}(-\mathbf{b}^T \boldsymbol{\Lambda}^{-1} \mathbf{b}) = \mathbf{0}$$

$$\mathbf{x}^g = \overline{\mathbf{x}}^g - \left(\boldsymbol{\Phi}^g \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Phi}^g)^T\right)^{-1} \left(\boldsymbol{\Phi}^g \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Phi}^f)^T\right) (\mathbf{x}^f - \overline{\mathbf{x}}^f)$$

## Incremental inclusion of landmarks



# Feature matching algorithm

Start from a set of candidates for each landmark

**for** (all 4-tuple combinations of landmarks and candidates $\mathbf{x}_4$) **do**

  Initialize $\mathbf{x}^f = \mathbf{x}_4$

  Infer $\hat{\mathbf{x}}^g$ using (11) or (16), obtaining $\hat{\mathbf{x}}$

  **while** ($\hat{\mathbf{x}}$ fulfills the constraints in (9)) **do**

    **for** (all other landmarks, $\ell_k \notin \mathbf{x}^f$) **do**

      **for** (all candidates $\mathbf{c}_k$ for landmark $\ell_k$) **do**

        Add the candidate $\mathbf{c}_k$ to $\mathbf{x}^f$ to obtain $\mathbf{x}^f_{test}$

        Infer $\hat{\mathbf{x}}^g_{test}$ from $\mathbf{x}^f_{test}$ to obtain $\hat{\mathbf{x}}_{test}$

        Compute the resulting cost $\gamma(\mathbf{c}_k)$ as in (17)

      **end for**

      Compute the landmark cost $\gamma(k) = \min_k \gamma(\mathbf{c}_k)$

    **end for**

    Update $\mathbf{x}^f$ adding the landmark with minimum $\gamma(k)$

    Infer $\hat{\mathbf{x}}^g$ from the updated $\mathbf{x}^f$ to obtain $\hat{\mathbf{x}}$

  **end while**

  Compute the score for $\mathbf{x}_4$ as $\#(\mathbf{x}^f) + e^{-\gamma(k)}$

**end for**

Keep the subset that achieved the highest score

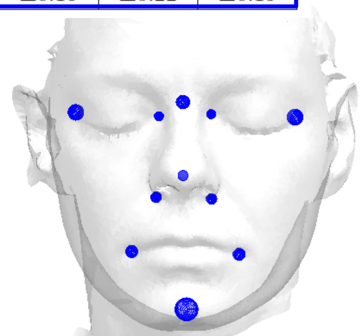
$$\gamma(\mathbf{c}_k) = median(\Delta\hat{\mathbf{x}}_{test})$$

$$\Delta\hat{\mathbf{x}}_{test} = \begin{cases} \|\hat{\mathbf{x}}_{test}(\ell_j) - \mathbf{x}^f_{test}(\ell_j)\|^2, & \forall \ell_j \in \mathbf{x}^f_{test} \\ \min_{\mathbf{c}_j} \|\hat{\mathbf{x}}_{test}(\ell_j) - \mathbf{c}_j\|^2, & \forall \ell_j \notin \mathbf{x}^f_{test} \end{cases}$$

# Results

| Landmark | n | prn | pg | en | ex | ac | ch |
|---|---|---|---|---|---|---|---|
| Passalis et al. [14] | n/a | 2.89(*) ±0.15 | 9.19(*) ±0.97 | 3.42 ±0.66 | 6.98(*) ±1.35 | n/a | 5.88(*) ±0.96 |
| Segundo et al. [4] | n/a | 2.63 ±0.13 | n/a | 5.64(*) ±0.61 | n/a | 4.93(*) ±0.21 | n/a |
| SRILF | 3.08 ±0.22 | 2.43 ±0.15 | 4.52 ±0.25 | 2.26 ±0.20 | 3.67 ±0.18 | 2.45 ±0.22 | 2.69 ±0.19 |

- Dataset of healthy volunteers (144 facial scans)
- 6-fold cross validation
- 11 facial landmarks
- Mean +/- standard error [mm]
- Significantly lower errors than the alternative methods compared

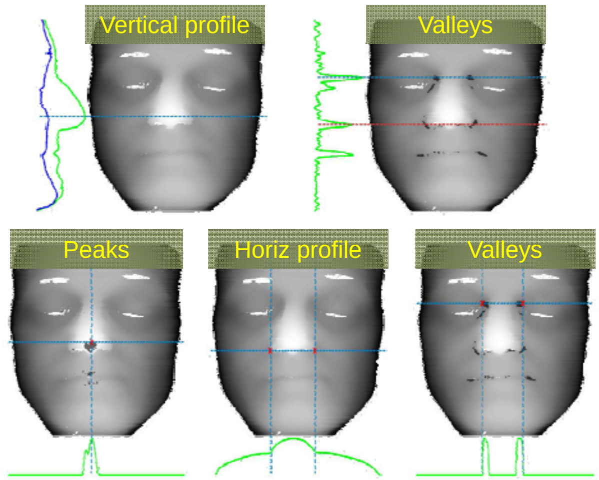


The radius of the spheres equals the average localization error

## Ad-hoc rules to locate landmarks

Segundo, M., et al. (2010). Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 40(5):1319–1330.

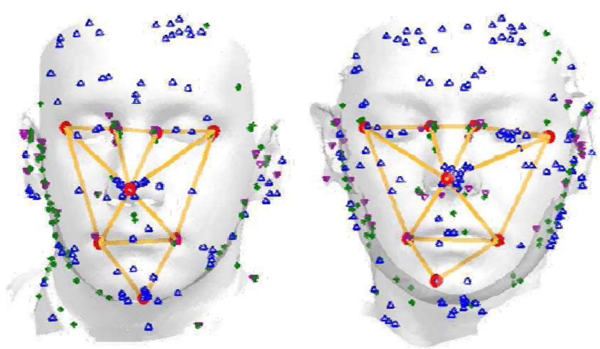
- Combining basic features (e.g. curvature, profile projections) with heuristic rules.
- Problems:
  - Scalability (to other landmarks),
  - Interdependency of rules
  - Orientation-dependant



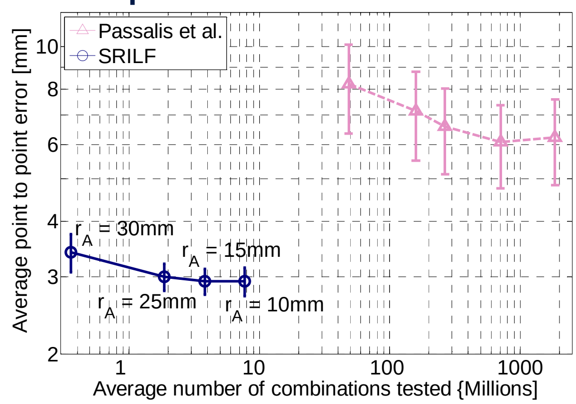

## Global geometric constraints

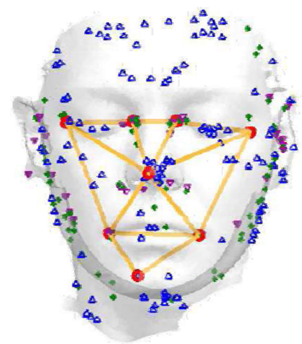
Passalis, G., et al. (2011). Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1938–1951.

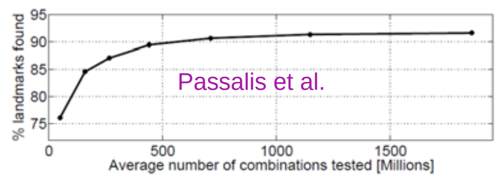
- Keep the top-N candidates for each landmark and test <u>all possible combinations</u>
  - Use statistical constraints to validate combinations
- Problems
  - Up to billions of combinations to test for just 8 landmarks
  - High computational load
  - High chance of accepting wrong combinations



# Comparison to Passalis et al.

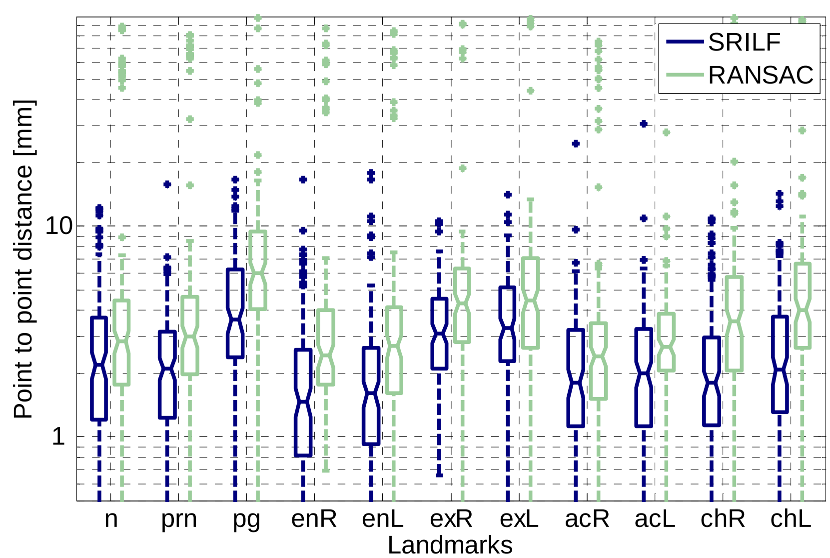





The method by Passalis et al. was unable to locate the landmarks for all meshes in our dataset

# Comparison to a rigid model



## Conclusions & further work

- We achieved an average accuracy of 3.2 mm targeting 11 facial landmarks
  - Results compare favourably to state of the art methods
  - The use of a flexible model performed significantly better than the rigid-model alternative
- The chin tip and outer-eye corners proved the most difficult within the addressed group
- We found that a key limitation is the local accuracy of spin images
  - Experiments using different descriptors indicate that localization errors may be further reduced by 10% – 20%

F.M. Sukno, J.L. Waddington and P.F. Whelan. Comparing 3D Descriptors for Local Search of Craniofacial Landmarks. ISVC 2012, pp 92-103.

## The Face3D project

Supported by wellcome trust

The project is funded by the Wellcome Trust

The partners in the project are:

- The University of Glasgow
- Royal College of Surgeons in Ireland
- Dublin City University
- Institute of Technology, Tralee
- University of Limerick

Face3D RESEARCH CONSORTIUM

MARIE CURIE ACTIONS

THANK YOU FOR YOUR ATTENTION