Automatic 3D Facial Region Retrieval from Multi-pose Facial Datasets

Theoharis Theoharis

University of Athens

- P. Perakis, T. Theoharis, G. Passalis and I. A. Kakadiaris

Motivation & Challenges

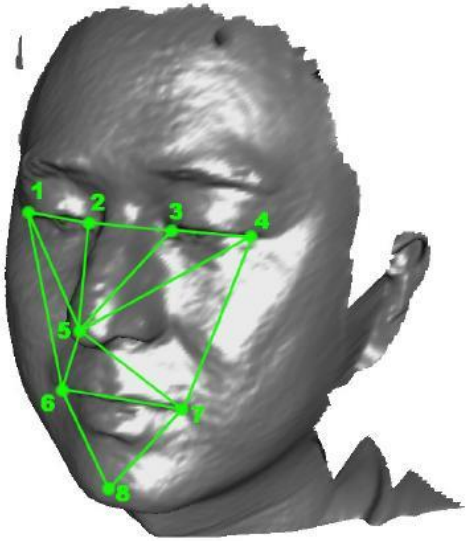
- Motivation
  - Growing size and importance of 3D object databases
  - Retrieval of specific facial areas (e.g., nasal region)
- 3D Landmark Detection
- 3D Landmark Model
- 3D Landmark Identification
- Challenges
  - Pose invariant method
  - Fully automatic method
  - Design a *robust and pose invariant* facial region detection and retrieval system, that works in datasets with pose rotations of up to 80° around the y-axis

Overview of Our Approach

- Training phase
  - Creation of 3D Landmark Models
    - Mean Landmark Shape (Procrustes Analysis)
    - Landmark Shape Variations (Principal Component Analysis)
  - Selection of Spin Image Templates
- Retrieval phase
  - Landmark Detection & Identification
    - Extract candidate Landmarks from Shape Index extrema
    - Classify Landmarks by using Spin Image templates
    - Select and identify Landmarks by fitting them with the Landmark Models
  - Facial Region Retrieval

The Active Landmark Models

- Use of 8 anatomical landmarks
- Creation of 3 Active Landmark Models (ALMs)
  - ALM8 (Landmarks: 1,2,3,4,5,6,7,8): Visible on frontal datasets
  - ALM5L (Landmarks: 1,2,5,6,8): Visible on left facial datasets
  - ALM5R (Landmarks: 3,4,5,7,8): Visible on right facial datasets



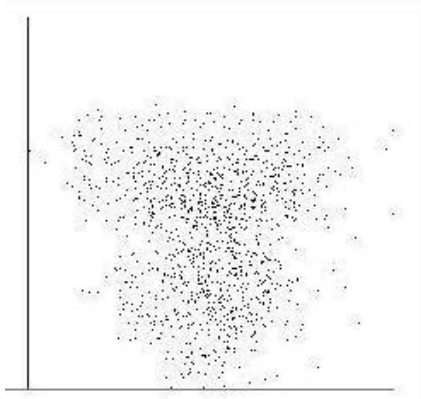
1. Right eye outer corner
2. Right eye inner corner
3. Left eye inner corner
4. Left eye outer corner
5. Nose tip
6. Mouth right corner
7. Mouth left corner
8. Chin tip

Creation of ALMs

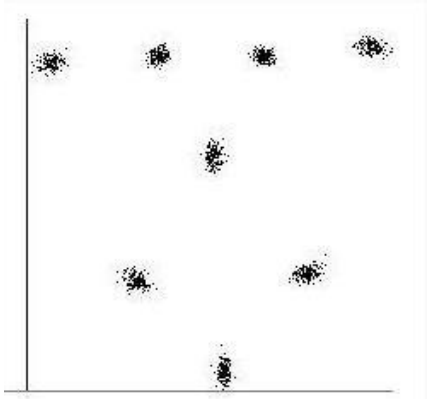
- Calculate the statistical Mean Shape for each landmark set (ALM8, ALM5L, ALM5R)
  - **ALM**s were calculated from a manually annotated training set of 150 frontal facial data sets with neutral expressions from FRGC v2, using **Procrustes Analysis**.
- Calculate the variations of each ALM
  - Landmark shape variations were calculated by applying Principal Component Analysis (**PCA**) to the aligned training set.
  - **Principal eigenvalues** were selected so that the model represents a given proportion (e.g.,99%) of the total variance of the landmark positions.
  - Principal eigenvalues correspond to the **modes of variations**:

$$\left|b_i\right| \leq 3\sqrt{\lambda_i}$$

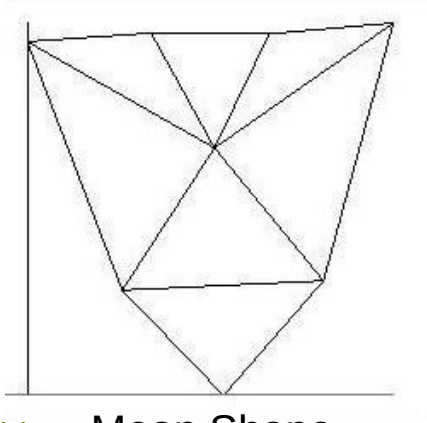
Mean Landmark Shape (ALM8)



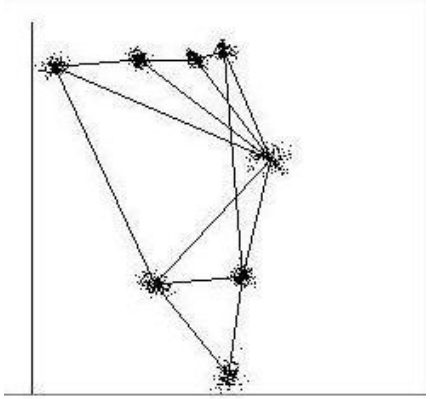
(a) Unaligned Landmarks



(b) Aligned Landmarks



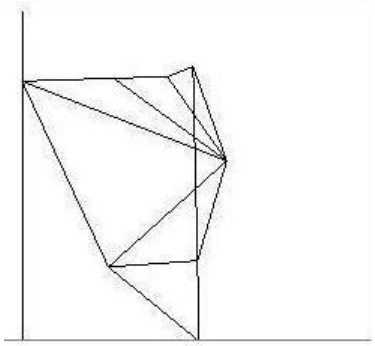
(c) Mean Shape



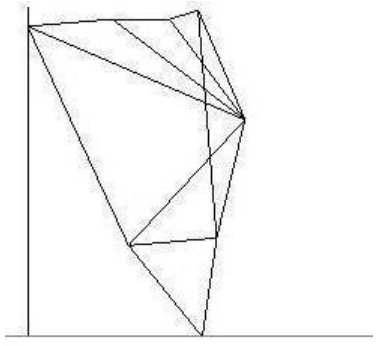
(d) Landmark Clouds & Mean Shape
(rotated 60 around the y-axis)

Landmark Shape Variations (ALM8)

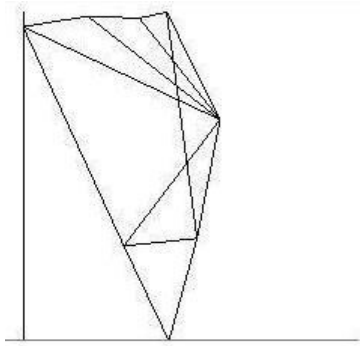
(a) 1st Mode of Variations (expresses face shape: Round vs Oval)



$b_1 = -3\sqrt{\lambda_1}$

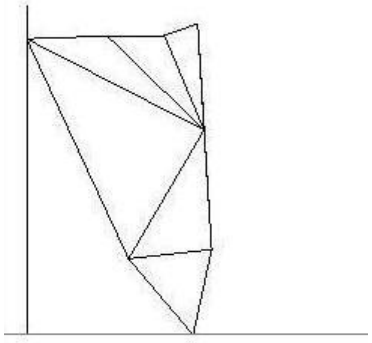


$b_1 = 0$

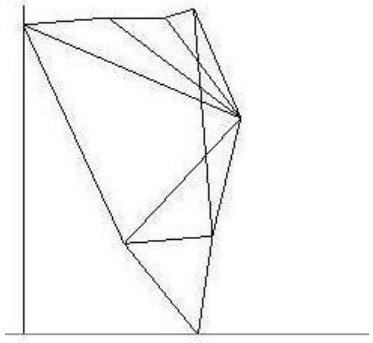


$b_1 = +3\sqrt{\lambda_1}$

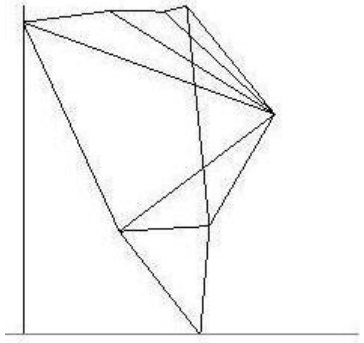
(b) 2nd Mode of Variations (expresses nose shape: Flat vs Extruded)



$b_2 = -3\sqrt{\lambda_2}$



$b_2 = 0$



$b_2 = +3\sqrt{\lambda_2}$

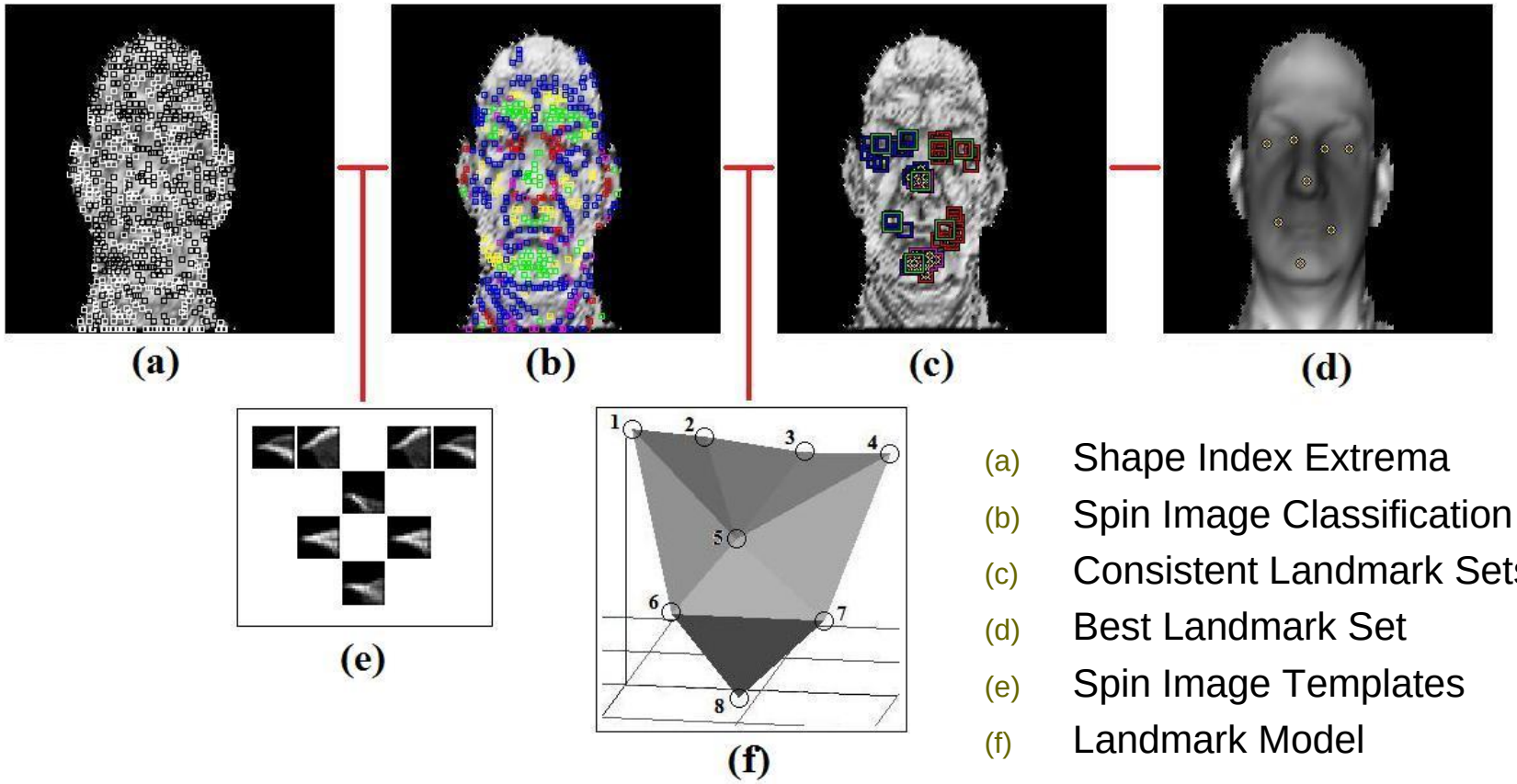
Landmarks Identification

- Landmark detection methods are not able to **identify** and **label** the detected landmarks
- To address this problem we use the ALMs
- **Labeling** is done by fitting a candidate landmark set to the corresponding ALM
- **Fitting** is done by minimizing the Procrustes distance of a candidate landmark set from the corresponding ALM by applying a rigid transformation
- A landmark set is considered as a **plausible shape** by checking the deformation parameters **b** to be within certain margins

Landmark Detection Procedure

- Landmark Detection & Identification
  - Extract candidate Landmarks from Shape Index maps
  - Classify candidate Landmarks by matching them with the corresponding Spin Image Templates
  - Compute the rigid transformation that best aligns combinations of 5 and 8 landmarks with a Landmark Model (ALM5L, ALM5R, ALM8)
  - Select the combination of Landmarks that has the minimum Procrustes distance from the ALMs
  - Discard combinations of candidate Landmarks that are not consistent with the ALMs
  - Label Landmarks and estimate pose

Landmark Detection Procedure



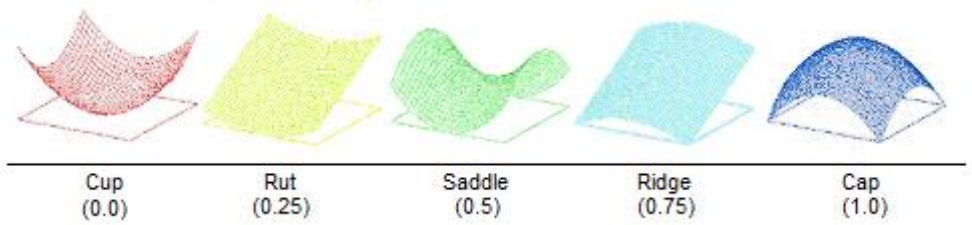
(a) Shape Index Extrema
(b) Spin Image Classification
(c) Consistent Landmark Sets
(d) Best Landmark Set
(e) Spin Image Templates
(f) Landmark Model

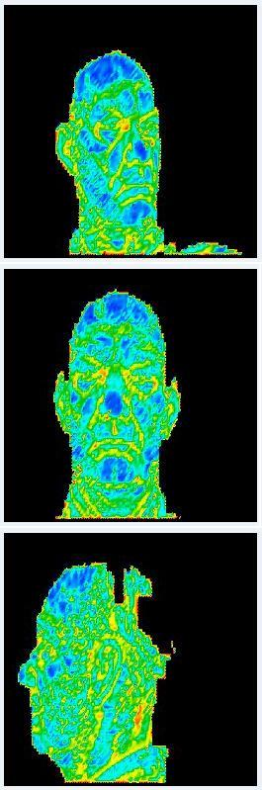
Shape Index extrema

- ***Shape Index*** represents the type of local curvature of a 3D object at a point **p**.

$$SI(p) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{k_{max}(p) + k_{min}(p)}{k_{max}(p) - k_{min}(p)}$$

- To find candidate Landmarks:
  - Create Shape Index maps
  - Local maxima (Caps) are candidate landmarks for nose and chin tips
  - Local minima (Cups) are candidate landmarks for eye and mouth corners

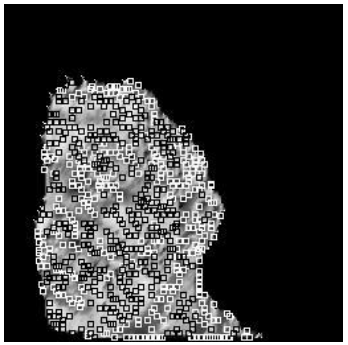




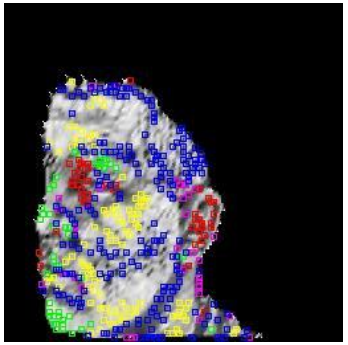
Shape Index maps of facial datasets

Spin Image classification

- ***Spin Image*** at a point **p** is a 2D grid accumulator of 3D points, as the grid is rotated around the normal **n** by 360°.
- It is a signature of the shape of an object at the neighborhood of **p**.
- To classify candidate Landmarks:
  - Shape Index's local maxima and minima are further classified into 5 classes according to their similarity with the 5 **Spin Image Templates**.
    - Red: eye inner corner
    - Blue: eye outer corner
    - Green: nose tip
    - Purple: mouth corner
    - Yellow: chin tip

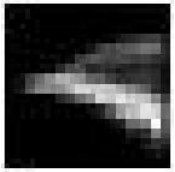


(a) Shape Index's maxima & minima

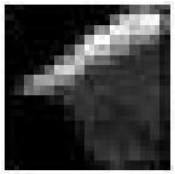


(b) Spin Image classification

Spin Image Templates

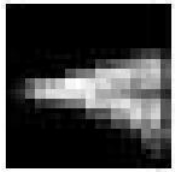


(a) Eye outer corner



(b) Eye inner corner

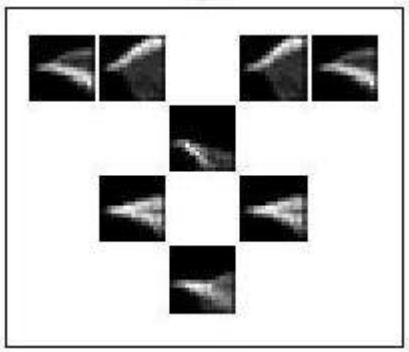
(c) Nose tip



(d) Mouth corner

(e) Chin tip

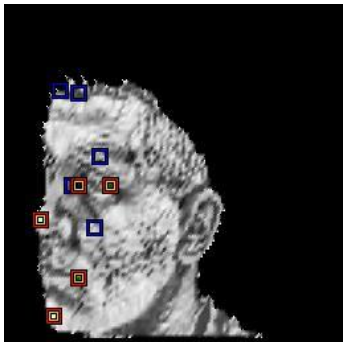
- **Spin Image Templates** were manually selected from facial datasets to best represent Spin Images at Landmark positions.
- Spin Image size:
  - Grid size: 16x16 bins
  - Bin size: 2mm
  - Spanned area: 3.2cm x 3.2cm



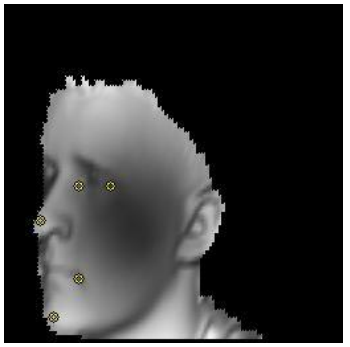
Spin Images Model

Resulting Landmarks

- From the classified candidate landmark points we create combinations of **5 landmarks**.
- The ones that do not conform with neither **ALM5L** nor **ALM5R** are filtered out.
- Finally we fuse them in complete landmark sets of **8 landmarks**.
- The ones that do not conform with **ALM8** are discarded.
- The **resulting Landmark set** is the one that has the minimum Procrustes distance to the corresponding model.



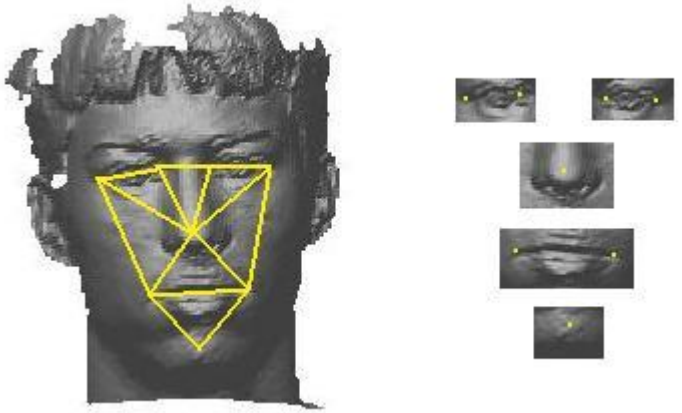
(a) Best Landmark sets



(b) Resulting Landmarks

Facial Region Retrieval

- After landmark localization and pose estimation is performed, facial region of interest is retrieved, by a 3D clipping procedure against a suitable **clipping volume**.
  - The position, orientation and size of the clipping volume is determined according to the landmark positions on the facial dataset.



(a) Landmark Localization

(b) Face Segmentation and Retrieval

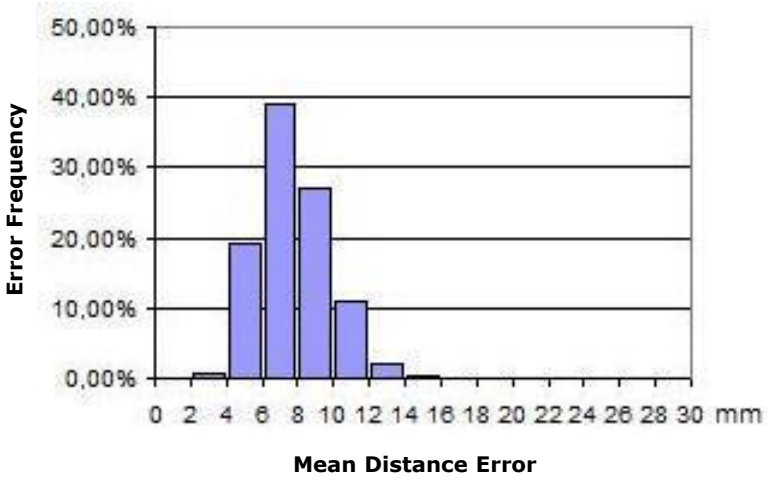
Test Databases

- Datasets from FRGC v2 Database
  - **DB00F**
    - 975 manually annotated frontal facial datasets
    - Contains several subjects with facial expressions
- Datasets from UND Database
  - **DB45L**
    - 115 manually annotated facial datasets at 45° left
  - **DB45R**
    - 115 manually annotated facial datasets at 45° right
  - **DB60L**
    - 80 manually annotated facial datasets at 60° left
  - **DB60R**
    - 80 manually annotated facial datasets at 60° right

Performance Evaluation

- Results for **DB00F**

Correct Pose Estimation	966 / 975	99.08%
Distance Error (mm)	Mean	Std.Dev.
Right Eye Outer Corner	8.13	3.79
Right Eye Inner Corner	7.02	3.18
Left Eye Inner Corner	7.46	3.07
Left Eye Outer Corner	9.21	4.25
Nose Tip	5.23	3.28
Mouth Right Corner	8.30	4.53
Mouth Left Corner	9.83	4.47
Chin Tip	6.71	4.32
Mean Distance Error	7.74	2.05

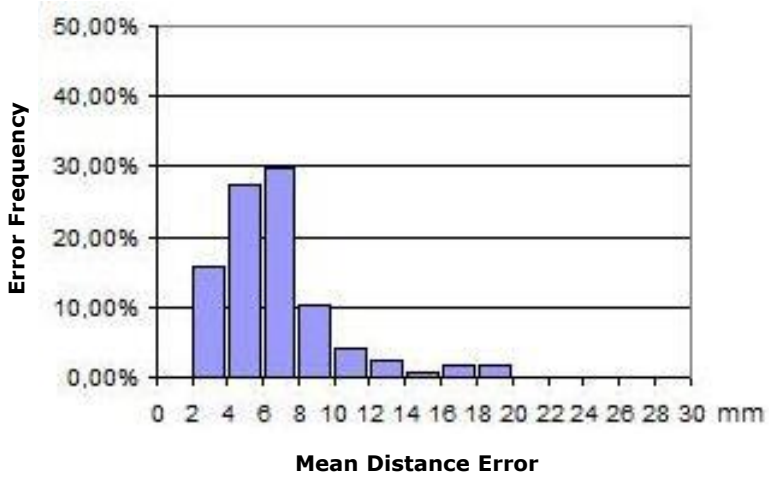


- Best located feature: Nose Tip
- Mean Error < 10.0mm: 86.1%

Performance Evaluation

- Results for **DB45R**

Correct Pose Estimation	114 / 115	99.13%
Distance Error (mm)	Mean	Std.Dev.
Right Eye Outer Corner	10.83	16.98
Right Eye Inner Corner	10.44	17.22
Nose Tip	9.93	18.11
Mouth Right Corner	11.09	17.79
Chin Tip	10.42	16.84
Mean Distance Error	10.54	16.84

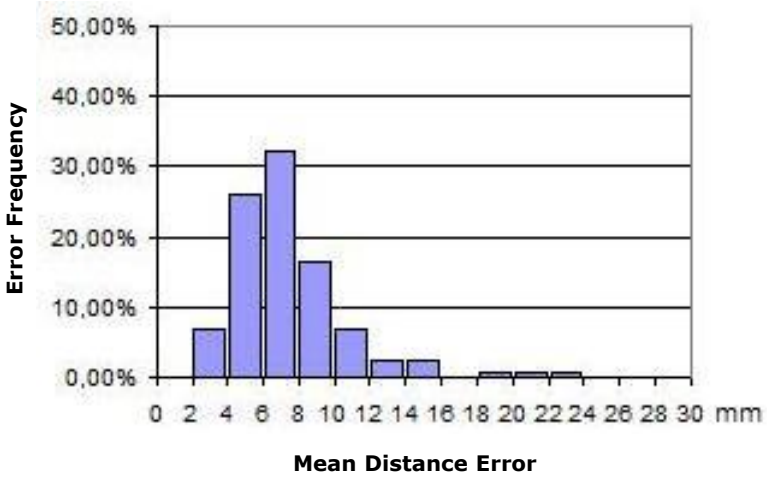


- Best located feature: Nose Tip
- Mean Error < 10.0mm: 83.3%

Performance Evaluation

- Results for **DB45L**

Correct Pose Estimation	115 / 115	100.00%
Distance Error (mm)	Mean	Std.Dev.
Left Eye Outer Corner	9.73	11.69
Left Eye Inner Corner	9.17	12.67
Nose Tip	8.29	13.18
Mouth Left Corner	11.03	13.83
Chin Tip	10.44	12.14
Mean Distance Error	9.73	12.09

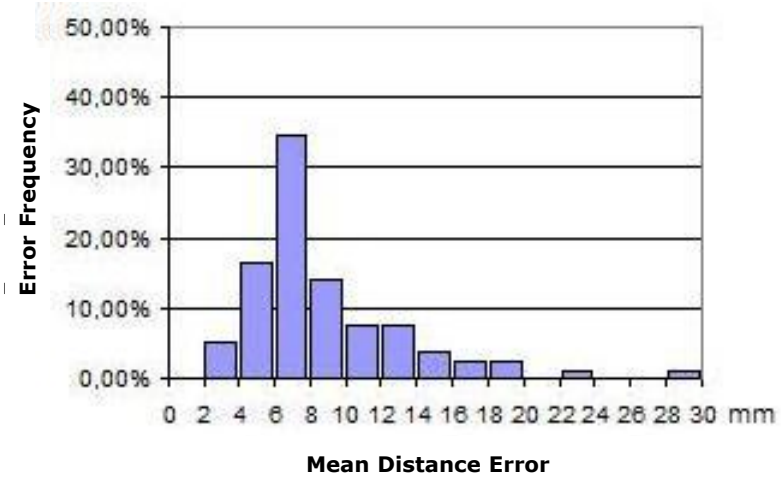


- Best located feature: Nose Tip
- Mean Error < 10.0mm: 81.8%

Performance Evaluation

- Results for **DB60R**

Correct Pose Estimation	78 / 80	97.50%
Distance Error (mm)	Mean	Std.Dev.
Right Eye Outer Corner	10.67	11.06
Right Eye Inner Corner	10.57	10.17
Nose Tip	8.73	10.35
Mouth Right Corner	11.51	10.65
Chin Tip	9.32	8.77
Mean Distance Error	10.16	9.14

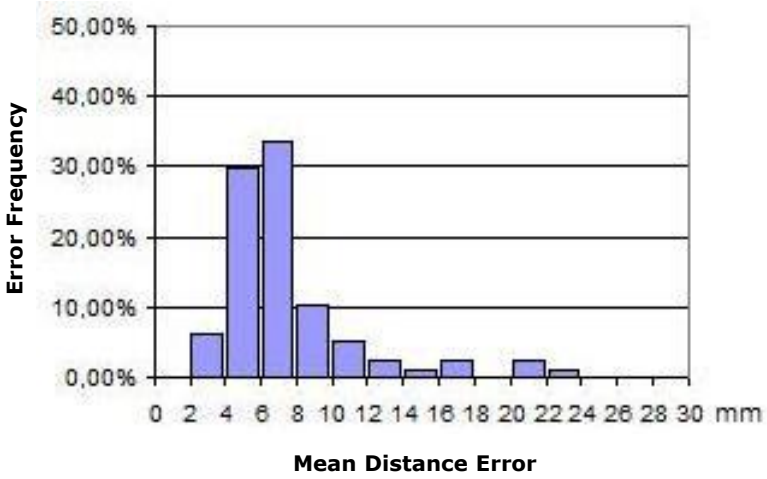


- Best located feature: Nose Tip
- Mean Error < 10.0mm: 70.5%

Performance Evaluation

- Results for **DB60L**

Correct Pose Estimation	77 / 80	96.25%
Distance Error (mm)	Mean	Std.Dev.
Left Eye Outer Corner	11.58	16.43
Left Eye Inner Corner	9.60	18.55
Nose Tip	9.57	18.52
Mouth Left Corner	11.43	18.02
Chin Tip	10.76	16.08
Mean Distance Error	10.59	17.06



- Best located feature: Nose Tip
- Mean Error < 10.0mm: 80.5%

Performance Evaluation

- General results:
  - Mean Error < 10.6 mm (irrespectively of pose)
  - Larger std. dev. on side datasets
  - Most robust feature:
    - the nose tip
  - Least robust feature:
    - the mouth corners
  - Correct estimation of pose: > 96%
  - Best performance:
    - frontal datasets (mean error: 7.74 mm)
  - Worst performance:
    - 60° left datasets (mean error: 10.59 mm)

Conclusion

- 3D facial region retrieval method:
  - Automatic,
  - Pose invariant,
  - Expression invariant,
  - Robust,

  even if half of the face is missing.

- Robustness enhancement:
  - Increase the ALM training examples (e.g., include expressions)
  - Increase the Landmark set (e.g., include nose base)
  - Statistically trained Spin Image Templates