1. Outline - Contributions

- Unsupervised detection of facial events
- Head pose, local actions of eyes, mouth, eyebrows etc.
- Classification of the extreme states; not precise calculation
- Important in Sign Language comprehension and recognition
- Lack of annotations. Manual Annotations are expensive

2. AAM fitting initialization

- AAM fitting estimates the shape and texture parameters vector \( \mathbf{q} \) that minimizes the error between the reconstructed textures and the image texture:

\[
\mathbf{q} = \arg\min_{\mathbf{q}} \sum_{i=1}^{N} \left( \mathbf{T}_i - \mathbf{S} + \mathbf{R} \mathbf{T}_i \right)^2
\]

- AAM fitting initialization framework:
  - High pose variation in Sign-Language videos
  - Need for robust and accurate AAM fitting
  - Initialization on each new video frame; no dynamics
  - Non-occlusion frame; detection from number of skin regions

- AAM fitting initialization framework:
  - High pose variation in Sign-Language videos
  - Need for robust and accurate AAM fitting
  - Initialization on each new video frame; no dynamics
  - Non-occlusion frame; detection from number of skin regions

3. Fitting and tracking results

- Comparison between proposed initialization and Viola-Jones face detection initialization framework
  - 76.7% MSE decrease

- Illustration of AAM similarity transform parameters
  - Face detection, skin detection, morphological operators

- Unsupervised method for Extreme States Classification (UnESC)

4. Local AAMs

- Model a specific facial area (mouth, eyes, brows etc.)
- Overcome the area's variance from the rest of the face
- Projection of Global AAM's fitting parameters to the eigenspace of the Local AAM

5. UnESC feature selection

- Possible features:
  - Global AAM eigenvectors parameters
  - Geometrical measures on the mask’s landmark points
  - Single-dimensional (1D) feature space
  - Facial event variation continuity set feature value change

6. Experimental results

I) Qualitative results (GSL)

- UnESC vs. Supervised Classification vs. Kmeans for pose yaw
- 729 experiments for various values of SPThres
- Supervised & Kmeans training set size 1+UnESC clusters size

II) Quantitative results (BU)

- UnESC vs. Supervised classification vs. Kmeans for pose yaw
- 729 experiments for various values of SPThres
- Supervised & Kmeans training set size 1+UnESC clusters size

7. 5. UnESC training

- Main idea: Partition 1D feature space in 3 representative clusters
- Two on the edges corresponding to the extreme states
- One on the center corresponding to the neutral state

- Step 1: Selection of 1D feature that best describes the facial event

- Step 2: Hierarchical Breakdown for density equalization
- Affine-invariant Hierarchical Clustering

- Step 3: Minimum-Distance Cluster Selection with Subjective Perceived Threshold (SPThres) parameter
- Some training data points are not classified in any cluster!

- Step 4: Training of 3-D Gaussian, one per cluster

8. Acknowledgments

This research work was supported by the EU under the project DeafCity with grant FP7ICT-233735 and in part by the项目DBiLIS with grant FP7ICT-233735.

For further information

Please contact: antonakos@ece.ntua.gr, vassi.pitsikalis@gmail.com

http://cvsp.cs.ntua.gr

Relevant work

  - UnESC: Interpretation and Coding of Face Images Using Appearance Models.

- L. Ding, S. Baker, A. I. Matthews.

- A. I. Matthews, L. Ding.

- P. Corcoran, S. Baker, I. Matthews, L. Ding.

- D. Brieden, I. Matthews, L. Ding.

- A. I. Matthews, L. Ding, S. Baker.

  - UnESC: Interpretation and Coding of Face Images Using Appearance Models.

- L. Ding, A.M. Martinez, S. Baker, I. Matthews.

- I. Matthews, L. Ding, S. Baker.

- P. Corcoran, S. Baker, I. Matthews, L. Ding.

- D. Brieden, I. Matthews, L. Ding.

- C. Goldenstein, S. Stegmann, S. Baker, A. I. Matthews, L. Ding.

- I. Matthews, L. Ding, S. Baker.

- A. I. Matthews, L. Ding, S. Baker.

- P. Corcoran, S. Baker, I. Matthews, L. Ding.

- D. Brieden, I. Matthews, L. Ding.