

Model-Based Facial Expression Recognition

Pedro Martins ♦ Jorge Batista
 Institute of Systems and Robotics
 Department of Electrical and Computer Engineering, University of Coimbra, Portugal

Abstract

- A framework for automatic facial expression recognition combining Active Appearance Model (AAM) and Support Vector Machines (SVM) is proposed.
- Seven different expressions of several subjects, representing the neutral face and the facial emotions of happiness, sadness, surprise, anger, fear, and disgust were analyzed.
- The human face is described by the AAM model, projecting the appearance results into the hyperplane that maximizes class separability using a multiclass SVM that emphasize the different expression categories.



Active Appearance Models



Shape Model

$$x = (x_1, y_1, \dots, x_n, y_n)^T$$

- Generalized Procrustes Analysis



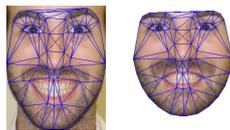
- Principal Components Analysis (PCA)

$$x = X + \Phi_s b_s$$

Texture Model

$$g = (g_1, g_2, \dots, g_{m-1}, g_m)^T$$

- Piecewise Affine Warp



- Low Memory PCA

$$g = \bar{g} + \Phi_g b_g$$

Combined Model

- Remove correlations between shape and texture model parameters

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s^{-1} \Phi_s^T (x - \bar{x}) \\ \Phi_g^T (x - \bar{x}) \end{pmatrix} \quad \Phi_c = \begin{pmatrix} \Phi_{cs} \\ \Phi_{cg} \end{pmatrix} \quad \rightarrow \quad x = \bar{x} + \Phi_s W_s^{-1} \Phi_{cs} c$$

$$g = \bar{g} + \Phi_g \Phi_{cg} c$$

- c is a vector of appearance controlling shape and texture

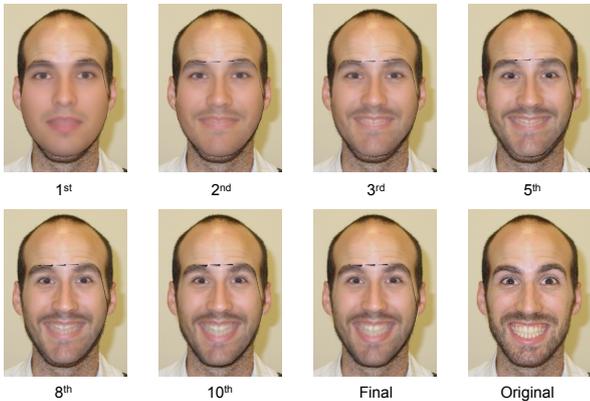
Model Training

$$\arg \min_c \sum_{p \in \text{test}} \left\| \begin{pmatrix} \text{Image} \\ \text{Pose} \end{pmatrix} - \begin{pmatrix} \text{Model} \\ \text{Pose} \end{pmatrix} \right\|^2$$

- Updating the appearance parameters, c, and pose $p = [c^T \quad ((s-1)\theta \quad T_x \quad T_y)^T]^T$
- The texture residual is defined as $r(p) = g_{image}(p) - g_{model}(p)$
- The goal is to find the optimal update, δp , to minimize $|r(p)|^2$
- Leading to $\delta p = -(J^T J)^{-1} J^T r$
- Training stage consists in estimating the Jacobian matrix, J

Model Fitting

- Model parameters are updated over texture residuals by $p_k = p_{k-1} - \alpha (J^T J)^{-1} J^T r$
- AdaBoost initial estimate for the location of the face



Facial Expression Recognition

- Describe a face using the AAM model and retrieve the appearance, c
- Project each appearance vector into the hyperspace that maximize class separability using a multiclass Support Vector Machines

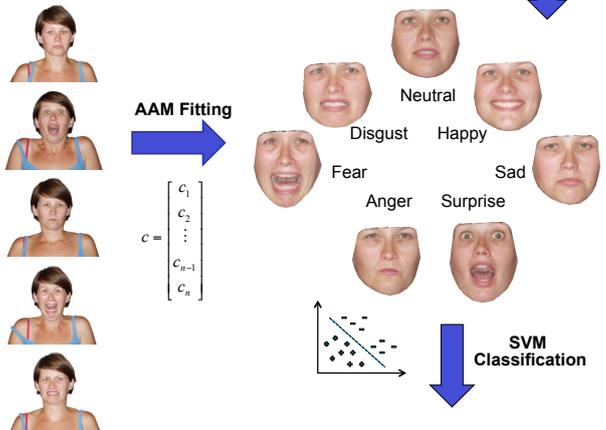
Facial Expression Database

- 147 Images
- 21 Individuals
- 7 Different Facial Expressions
- 640 x 480 Color Images

- How much variance should be held on the AAM ?

| Variance (%) | N° Combined EigenVectors |
|--------------|--------------------------|
| 95 | 17 |
| 97 | 29 |
| 98 | 42 |
| 99 | 70 |
| 99.5 | 97 |
| 99.9 | 133 |

Build Several Models



One-Against-All Support Vector Machines

- Facial expression SVM classification was achieved using a multiclass one-against-all voting scheme with a RFB kernel, $K(x_j, x_i) = e^{-\gamma \|x_j - x_i\|^2}$
- The kernel parameter, γ , and the misclassification penalty C were found by crossvalidation
- Feature data was normalized by mapping-it into the unitary hypersphere

Confusion Matrices – Leave-One-Out Cross-Validation

| | Neut | Happ | Sad | Surp | Ang | Fear | Disg |
|------|-------|------|-------|------|------|-------|------|
| Neut | 56.25 | 0 | 31.25 | 0 | 6.25 | 0 | 6.25 |
| Happ | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| Sad | 37.5 | 0 | 50 | 0 | 0 | 6.25 | 6.25 |
| Surp | 0 | 0 | 0 | 75 | 6.25 | 18.75 | 0 |
| Ang | 0 | 6.25 | 6.25 | 0 | 50 | 12.5 | 25 |
| Fear | 12.5 | 0 | 6.25 | 37.5 | 6.25 | 31.25 | 6.25 |
| Disg | 0 | 12.5 | 6.25 | 0 | 37.5 | 6.25 | 37.5 |

- Effect of confusion between the pairs of expressions

- Neutral/Sad
- Anger/Disgust

- Overall Recognition Rate = 57.14% (Dataset 95%)

| | Neut | Happ | Surp | Fear | Disg |
|------|------|-------|-------|-------|-------|
| Neut | 100 | 0 | 0 | 0 | 0 |
| Happ | 6.25 | 93.75 | 0 | 0 | 0 |
| Surp | 0 | 6.25 | 81.25 | 12.5 | 0 |
| Fear | 12.5 | 6.25 | 31.25 | 43.75 | 6.25 |
| Disg | 0 | 12.5 | 6.25 | 0 | 81.25 |

Removing Correlated Facial Expressions

- Overall Recognition Rate = 80% (Dataset 99.5%)

