Facial Expression Recognition Using Active Appearance

Pedro Martins
Joana Sampaio
Jorge Batista

Institute of Systems and Robotics
Dep. of Electrical Engineering and Computers
University of Coimbra

VISAPP

Introduction

- Facial expression has more influence than simple audio information
- Human Computer Interface (HCI)
- Video compression

- Recognition of 7 different expressions
  - [Ekman and Friesen] said that people are born with the ability to generate and interpret only six facial expressions: happiness, sadness, surprise, anger, fear and disgust.
- Facial expression recognition in still images

Agenda

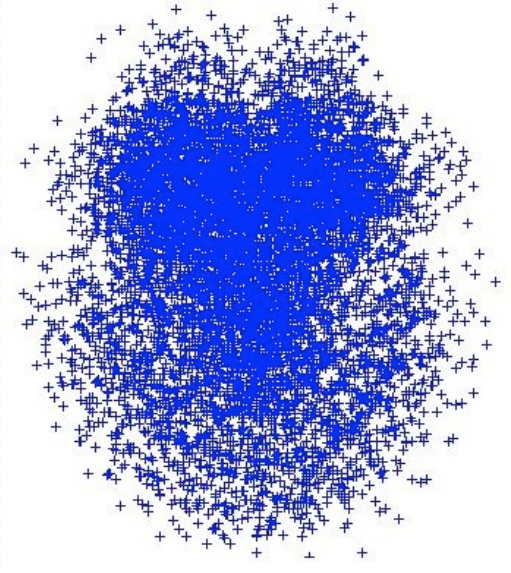
- Active Apperance Models
  - Shape Model
  - Texture Model
  - Combined Model
- Linear Discriminant Analysis
- Classification using *malahonobis* distance
- Results

Active Appearance Models

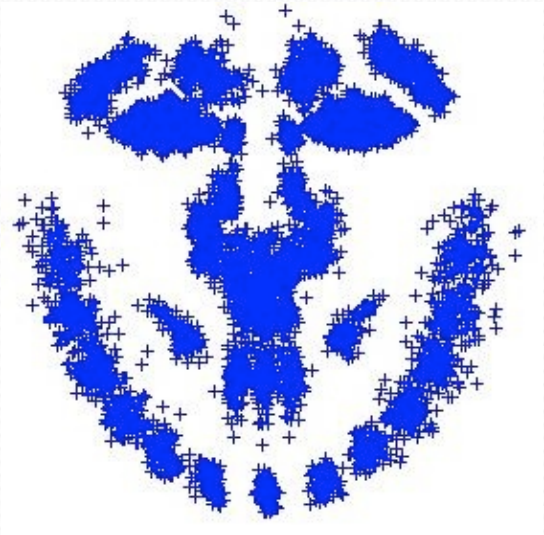
- AAM is a statistical based segmentation method, where the variability of shape and texture is captured from a dataset
- Able to extract relevant face information without background interference
- Describes facial characteristics in a reduced model

Shape Model

- Shape defined as $x = (x_1, y_1, \ldots, x_n, y_n)^T$
- Generalised Procrustes Analysis (GPA)
  - Remove Location, scale and rotation effects



Raw Data



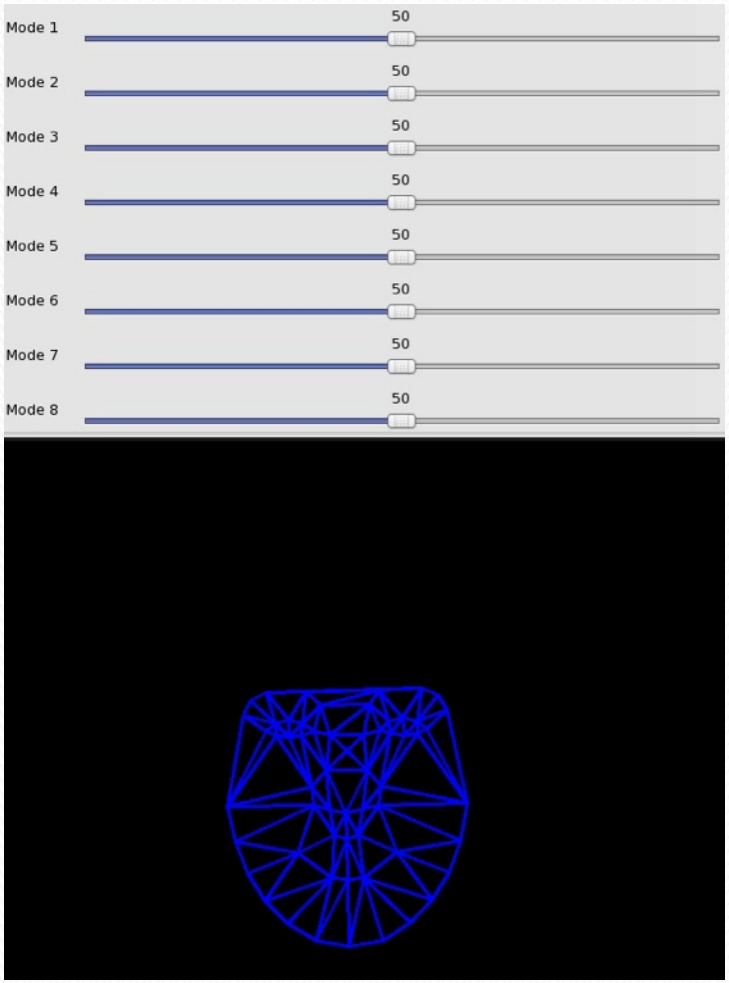
Aligned Data

Statistical Shape Model

- Applying a PCA:

$$x = \bar{x} + \Phi_s b_s$$

  - x is the synthesized
  - $\bar{x}$ *is the mean shape*
  - *Φs contains the highest covariance texture eigenvectors*
  - *bs* is a vector of shape parameters representing the weights

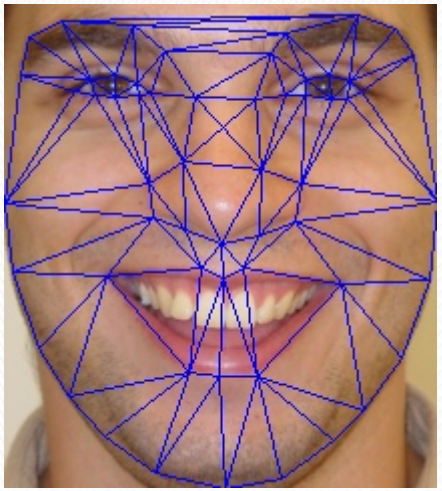


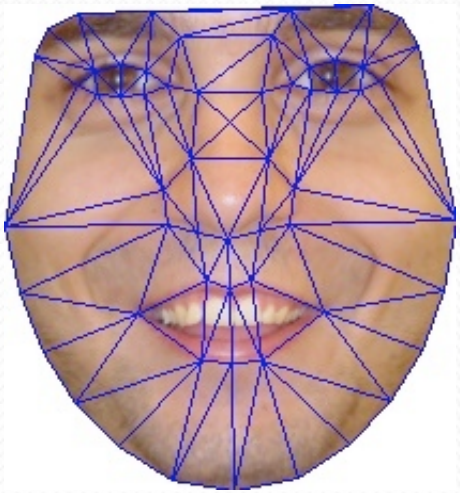
Texture Model

- For *m* pixels sampled, the texture is represented by:

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

- Required warping each image to a common reference frame





- Delaunay Triangulation
- Each pixel is mapped barycentric coordinates

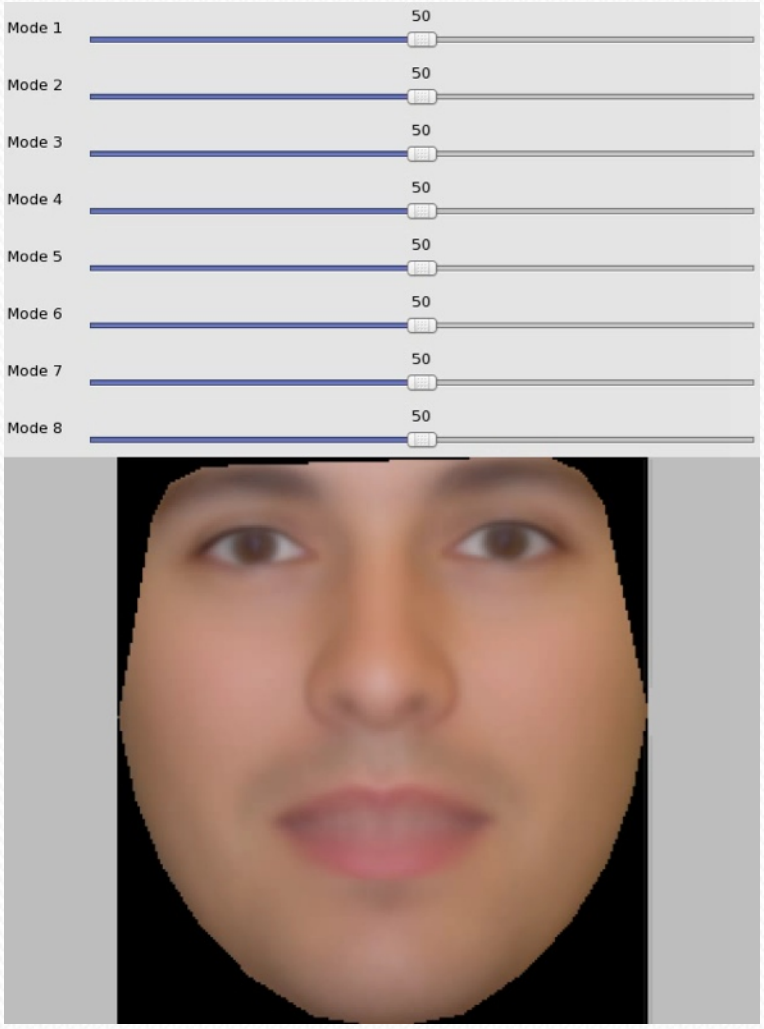
Texture Mapping Examples

Statistical Texture Model

- *Applying a LowMemory* PCA:

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

  - g is the synthesized texture
  - $\bar{g}$ is the mean texture
  - Φg contains the highest covariance texture eigenvectors
  - bg is a vector of texture parameters



Combined Model

- To remove correlations between bs and bg a third PCA is performed

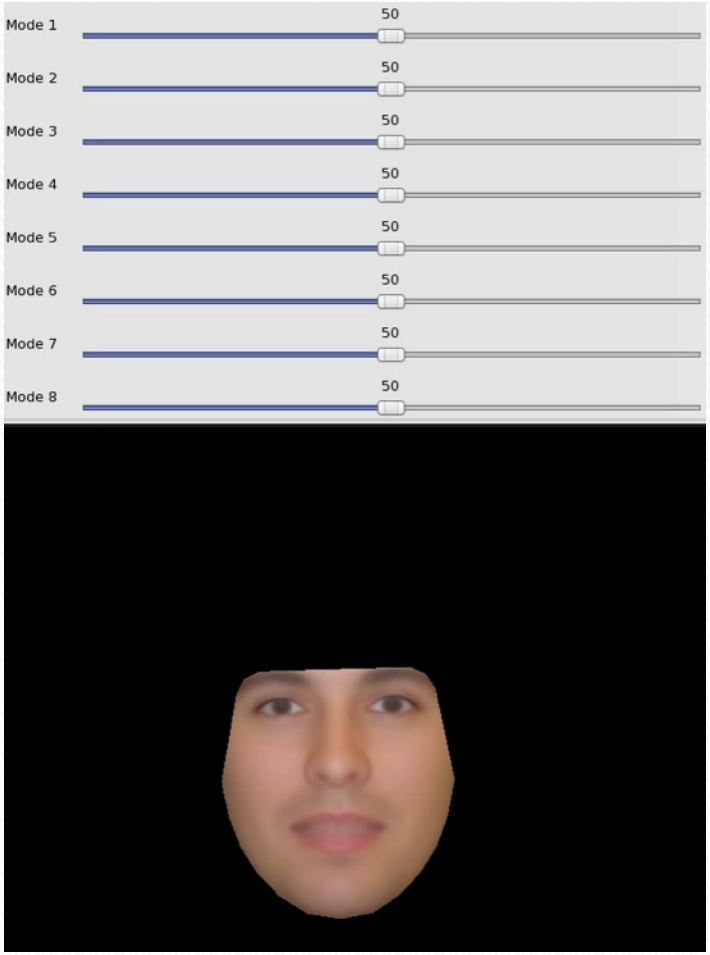
$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s \Phi_s^T (x - \bar{x}) \\ \Phi_g^T (g - \bar{g}) \end{pmatrix} \qquad W_s = rI \qquad r = \frac{\sum_i \lambda_{gi}}{\sum_j \lambda_{sj}}$$

$$b = \Phi_c c \qquad \Phi_c = \begin{pmatrix} \Phi_{cs} \\ \Phi_{cg} \end{pmatrix}$$

$$x = \bar{x} + \Phi_s W_s^{-1} \Phi_{cs} c \qquad g = \bar{g} + \Phi_g \Phi_{cg} c$$

- Building an AAM instance

Shape | Texture in the mean shape frame | AAM instance



AAM Instance Examples

Facial Expression Recognition

Disgust

Neutral

Happy

Fear

Sad

Anger

Surprise

AAM Retained Variance

- How much variance should be retained in AAM building process?

Variance (%)	Number of Combined EigenVectors
95	17
97	29
98	42
99	70
99.5	97
99.9	133

95%

99%

99.9%

Linear Discriminant Analysis

- Supervised method that maximizes the between-class variance as well that minimizes the within-class variance

- Classification using *malahanobis* distance

$$D = (c - \overline{c_i}) \sum^{-1} (c - \overline{c_i})$$

$$S_b = \sum_{i=1}^{n_c} n_i (\overline{x_i} - \overline{x})(\overline{x_i} - \overline{x})^T$$

Between-class scatter matrix

$$S_w = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} (x_{i,j} - \overline{x_i})(x_{i,j} - \overline{x_i})^T$$

Within-class scatter matrix

$x_{i,j}$ is the j^{th} sample in class i
$\overline{x_i}$ mean of class i
$\overline{x}$ mean of all classes
n_c number of classes
n_i number of samples in class i

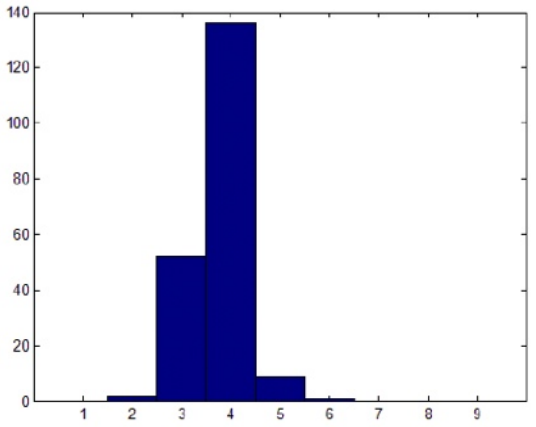
LDA Evaluation Metric

- How many eigenvectors hold on LDA?
  - k-means clustering on result of LDA
  - Discrimination quality given by:

$$DQuality = \sum_{i=1}^{n_{classes}} \max(row(i)) - \min(row(i))$$

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**N**	0	0	0	0	0	0
Happ	0	**N**	0	0	0	0	0
Sad	0	0	**N**	0	0	0	0
Surp	0	0	0	**N**	0	0	0
Ang	0	0	0	0	**N**	0	0
Fear	0	0	0	0	0	**N**	0
Disg	0	0	0	0	0	0	**N**

- K-means clustering result for ideal LDA



- Histogram for best LDA modes on 250 trials

Experimental Results 7 Expressions

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**19.0**	0	58.38	0	9.52	19.05	0
Happ	0	**90.48**	0	0	0	4.77	4.77
Sad	9.52	0	**61.90**	4.77	9.52	14.29	0
Surp	0	0	0	**80.95**	0	19.05	0
Ang	0	0	14.29	0	**33.33**	9.52	42.86
Fear	0	4.77	9.52	19.05	14.29	**52.38**	0
Disg	0	14.29	0	0	38.10	0	**47.62**

- Confusion Matrix 97% Overal Recognition Rate = 55%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**33.33**	0	61.90	0	0	4.76	0
Happ	0	**80.95**	0	0	0	9.52	9.52
Sad	0	4.76	**66.67**	0	19.05	9.52	0
Surp	0	0	0	**71.43**	0	28.57	0
Ang	0	4.76	4.76	0	**42.86**	14.29	33.33
Fear	0	4.76	9.52	19.05	9.52	**52.38**	4.76
Disg	0	9.52	0	0	38.10	4.76	**47.62**

- Confusion Matrix 98% Overal Recognition Rate = 56.5%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**52.38**	0	42.86	0	4.76	0	0
Happ	0	**90.48**	4.76	0	0	4.76	0
Sad	4.76	4.76	**76.19**	0	4.76	4.76	4.76
Surp	0	0	0	**76.19**	0	23.81	0
Ang	4.76	0	9.52	0	**33.33**	23.81	28.57
Fear	0	9.52	4.76	14.29	4.76	**66.67**	0
Disg	0	23.81	4.76	0	33.33	4.76	**33.33**

- Confusion Matrix 99% Overal Recognition Rate = 61.2%

- Results by a Leave-one-out cross-validation scheme

Experimental Results 7 Expressions

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**19.0**	0	58.38	0	9.52	19.05	0
Happ	0	**90.48**	0	0	0	4.77	4.77
Sad	9.52	0	**61.90**	4.77	9.52	14.29	0
Surp	0	0	0	**80.95**	0	19.05	0
Ang	0	0	14.29	0	**33.33**	9.52	42.86
Fear	0	4.77	9.52	19.05	14.29	**52.38**	0
Disg	0	14.29	0	0	38.10	0	**47.62**

- Confusion Matrix 97% Overal Recognition Rate = 55%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**33.33**	0	61.90	0	0	4.76	0
Happ	0	**80.95**	0	0	0	9.52	9.52
Sad	0	4.76	**66.67**	0	19.05	9.52	0
Surp	0	0	0	**71.43**	0	28.57	0
Ang	0	4.76	4.76	0	**42.86**	14.29	33.33
Fear	0	4.76	9.52	19.05	9.52	**52.38**	4.76
Disg	0	9.52	0	0	38.10	4.76	**47.62**

- Confusion Matrix 98% Overal Recognition Rate = 56.5%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**52.38**	0	42.86	0	4.76	0	0
Happ	0	**90.48**	4.76	0	0	4.76	0
Sad	4.76	4.76	**76.19**	0	4.76	4.76	4.76
Surp	0	0	0	**76.19**	0	23.81	0
Ang	4.76	0	9.52	0	**33.33**	23.81	28.57
Fear	0	9.52	4.76	14.29	4.76	**66.67**	0
Disg	0	23.81	4.76	0	33.33	4.76	**33.33**

- Confusion Matrix 99% Overal Recognition Rate = 61.2%

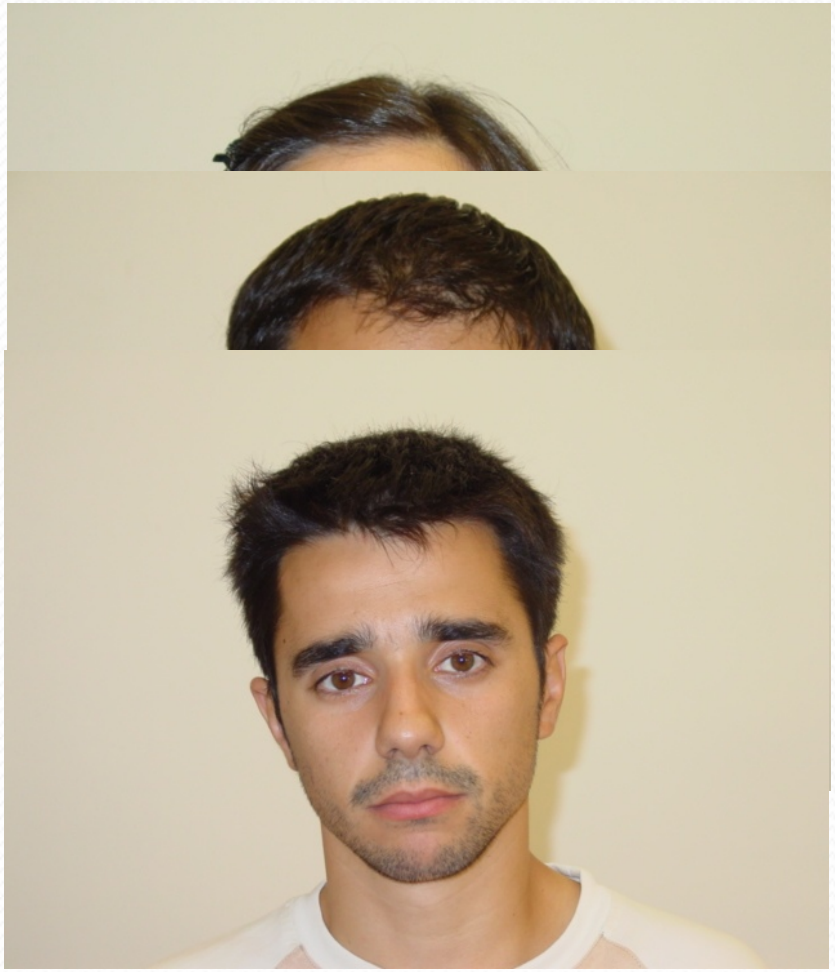
- Results by a Leave-one-out cross-validation scheme

Neutral/Sad Image Correlations

Neutral/Sad Image Correlations

Neutral/Sad Image Correlations

Neutral/Sad Image Correlations



Experimental Results 7 Expressions

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**19.0**	0	58.38	0	9.52	19.05	0
Happ	0	**90.48**	0	0	0	4.77	4.77
Sad	9.52	0	**61.90**	4.77	9.52	14.29	0
Surp	0	0	0	**80.95**	0	19.05	0
Ang	0	0	14.29	0	**33.33**	9.52	42.86
Fear	0	4.77	9.52	19.05	14.29	**52.38**	0
Disg	0	14.29	0	0	38.10	0	**47.62**

- Confusion Matrix 97% Overal Recognition Rate = 55%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**33.33**	0	61.90	0	0	4.76	0
Happ	0	**80.95**	0	0	0	9.52	9.52
Sad	0	4.76	**66.67**	0	19.05	9.52	0
Surp	0	0	0	**71.43**	0	28.57	0
Ang	0	4.76	4.76	0	**42.86**	14.29	33.33
Fear	0	4.76	9.52	19.05	9.52	**52.38**	4.76
Disg	0	9.52	0	0	38.10	4.76	**47.62**

- Confusion Matrix 98% Overal Recognition Rate = 56.5%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**52.38**	0	42.86	0	4.76	0	0
Happ	0	**90.48**	4.76	0	0	4.76	0
Sad	4.76	4.76	**76.19**	0	4.76	4.76	4.76
Surp	0	0	0	**76.19**	0	23.81	0
Ang	4.76	0	9.52	0	**33.33**	23.81	28.57
Fear	0	9.52	4.76	14.29	4.76	**66.67**	0
Disg	0	23.81	4.76	0	33.33	4.76	**33.33**

- Confusion Matrix 99% Overal Recognition Rate = 61.2%

Experimental Results 7 Expressions

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**19.0**	0	58.38	0	9.52	19.05	0
Happ	0	**90.48**	0	0	0	4.77	4.77
Sad	9.52	0	**61.90**	4.77	9.52	14.29	0
Surp	0	0	0	**80.95**	0	19.05	0
Ang	0	0	14.29	0	**33.33**	9.52	42.86
Fear	0	4.77	9.52	19.05	14.29	**52.38**	0
Disg	0	14.29	0	0	38.10	0	**47.62**

- Confusion Matrix 97% Overal Recognition Rate = 55%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**33.33**	0	61.90	0	0	4.76	0
Happ	0	**80.95**	0	0	0	9.52	9.52
Sad	0	4.76	**66.67**	0	19.05	9.52	0
Surp	0	0	0	**71.43**	0	28.57	0
Ang	0	4.76	4.76	0	**42.86**	14.29	33.33
Fear	0	4.76	9.52	19.05	9.52	**52.38**	4.76
Disg	0	9.52	0	0	38.10	4.76	**47.62**

- Confusion Matrix 98% Overal Recognition Rate = 56.5%

	Neut	Happ	Sad	Surp	Ang	Fear	Disg
Neut	**52.38**	0	42.86	0	4.76	0	0
Happ	0	**90.48**	4.76	0	0	4.76	0
Sad	4.76	4.76	**76.19**	0	4.76	4.76	4.76
Surp	0	0	0	**76.19**	0	23.81	0
Ang	4.76	0	9.52	0	**33.33**	23.81	28.57
Fear	0	9.52	4.76	14.29	4.76	**66.67**	0
Disg	0	23.81	4.76	0	33.33	4.76	**33.33**

- Confusion Matrix 99% Overal Recognition Rate = 61.2%

Anger/Disgust Image Correlations

Anger/Disgust Image Correlations

Anger/Disgust Image Correlations

Anger/Disgust Image Correlations

Facial Expression Recognition

Disgust

Neutral

Happy

Fear

Sad

Anger

Surprise

- Neuroscience studies have showned that, in order to percieve sad and anger expressions, an especific cognitive process is required [Killgore and Yurgelun-Todd]

Experimental Results 5 Expressions

	Neut	Happ	Surp	Fear	Disg
Neut	**76.19**	0	0	23.8	0
Happ	0	**95.23**	0	4.76	0
Surp	0	0	**61.9**	38.09	0
Fear	33.33	4.76	14.28	**42.85**	4.76
Disg	0	9.52	0	0	**90.47**

- Confusion Matrix 97% Overal Recognition Rate = 73.3%

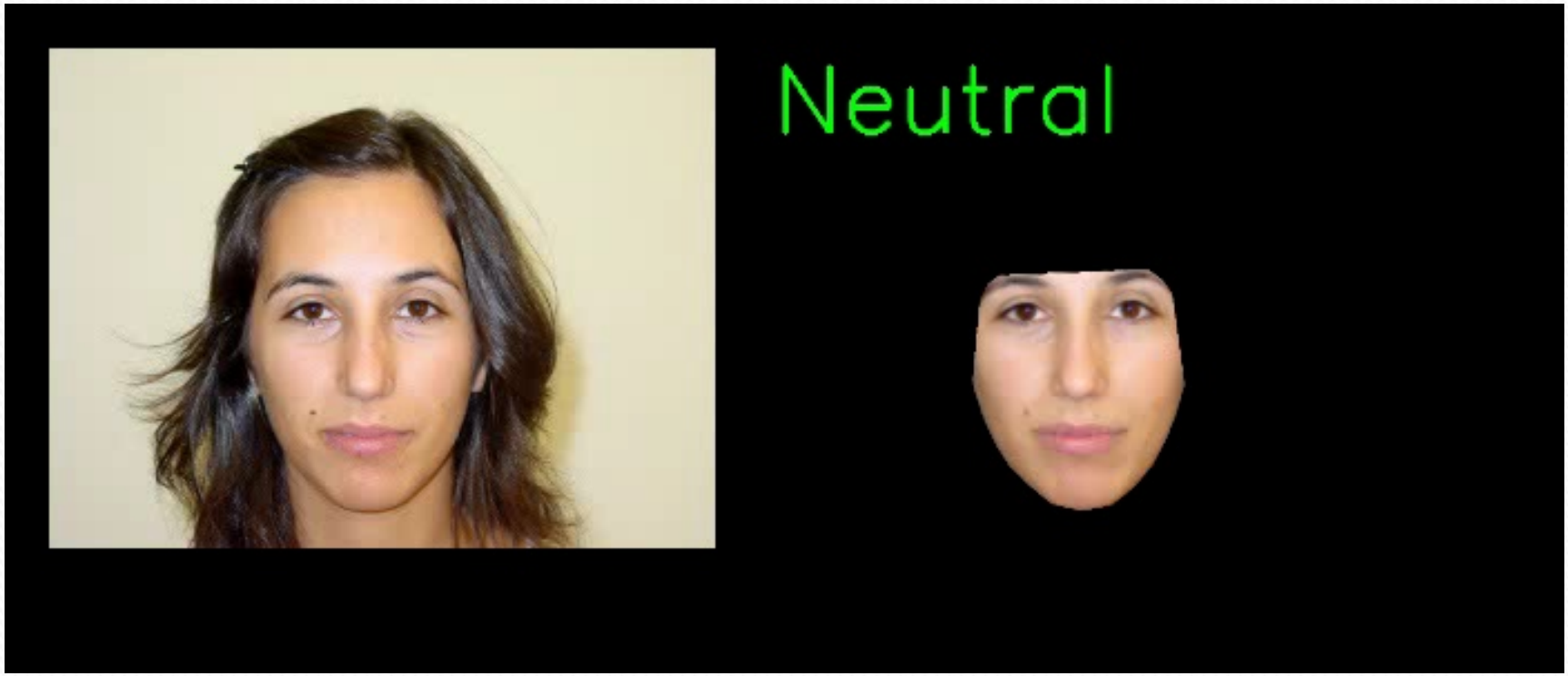
	Neut	Happ	Surp	Fear	Disg
Neut	**95.23**	0	0	4.76	0
Happ	0	**95.23**	0	0	4.76
Surp	9.52	0	**76.19**	14.28	0
Fear	19.04	4.76	19.04	**52.38**	4.76
Disg	0	19.04	4.76	4.76	**71.42**

- Confusion Matrix 98% Overal Recognition Rate = 78.09%

	Neut	Happ	Surp	Fear	Disg
Neut	**85.71**	4.76	4.76	4.76	0
Happ	9.52	**76.19**	0	4.76	9.52
Surp	0	0	**71.42**	28.57	0
Fear	4.76	9.52	33.33	**47.61**	4.76
Disg	9.52	14.28	4.76	0	**71.42**

- Confusion Matrix 99% Overal Recognition Rate = 70.47%

5 Expressions Classification Video



Final Notes

- Standart AAM to describe faces in a compact way
- LDA
  - Discrimination quality given by k-means
- Classification by mahalanobis distance
  - 7 Expressions: Best recognition rate=61.2% (with 99% AAM variance)
  - 5 Expressions: Best recognition rate=78.4% (with 98% AAM variance)

- Future work:
  - Classify expression in a video sequence using HMM