Model-Based Facial Expression Recognition

Pedro Martins, Jorge Batista
Institute for Systems and Robotics
Dep. of Electrical and Computer Engineering, FCT-University of Coimbra, Portugal
{pedromartins,batista}@isr.uc.pt

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 analysed. The proposed solution starts by describing the human face by an AAM model, projecting the appearance results to the hyperplane that maximizes class separability using a multiclass SVM that emphasize the different expression categories.

1. Introduction

Facial expression recognition plays an important role in human communication, having much more influence than just audio information. Psychology studies [2] show that there are six basic emotions universally recognized: joy, sadness, surprise, fear, anger and disgust. In this work we focus only on these facial expressions. Relevant facial information can be extracted from images using model-based approaches where statistical redundancy reduction techniques are used for compact coding, fully describing a face characteristics in a reduced model. Active Appearance Models (AAM) [5] is an effective way to locate facial features that models both shape and texture from an observed training set.

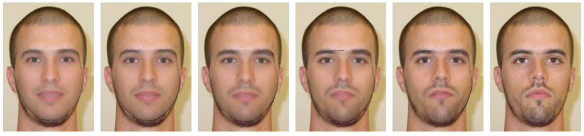
The proposed facial expression recognition system is based on the discriminating power of the AAM appearance parameters that is projected to the hyperplane that maximizes class separability using a multiclass Support Vector Machines (SVM) [1]. A series of experiments where conducted on a set of unknown images showing the faces of different subjects with facial expressions ranging neutral to intensely expressive.

2. Active Appearance Models (AAM)

AAM [5] is a statistical based template matching method, where the variability of shape and texture is captured from a representative training set. Principal Components Analysis (PCA) on shape and texture data allow to produce a parametrized face model that fully describe with photorealistic quality the trained faces as well as unseen. A shape is represented as n landmark points by a vector $\mathbf{x} = (x_1, y_1, \ldots, x_n, y_n)^T$. With several shape annotations follows a Procrustes analysis that removes similarity effects. Applying a PCA, the statistical variation can be modeled with $\mathbf{x} = \bar{\mathbf{x}} + \Phi_s \mathbf{b}_s$ where new shapes $\mathbf{x}$, are synthesized by deforming the mean shape, $\bar{\mathbf{x}}$, using a weighted linear combination of eigenvectors, Φ_s. $\mathbf{b}_s$ is a vector of parameters which represents the weights. The texture is represented by a vector $\mathbf{g}$ of m sampled pixels. Building such model requires warping each training image so that the control points match those of the mean shape. Delaunay triangles are establish to a piece-wise affine warp process, where each pixel is mapped into the destination triangle using barycentric coordinates with bilinear interpolation correction. The statistical texture model can be obtained by applying a low-memory PCA on these normalized textures, $\mathbf{g} = \bar{\mathbf{g}} + \Phi_g \mathbf{b}_g$, where $\mathbf{g}$ is the synthesized texture, $\bar{\mathbf{g}}$ is the mean texture, Φ_g contains the highest texture eigenvectors and $\mathbf{b}_g$ is a vector of parameters. In order to obtain further dimensionality reduction, the process of building an AAM also involves applying a futher PCA to the shape and texture parameters, that leads to $\mathbf{x} = \bar{\mathbf{x}} + \Phi_s \mathbf{W}_s^{-1} \Phi_{c,s} \mathbf{c}$ and $\mathbf{g} = \bar{\mathbf{g}} + \Phi_g \Phi_{c,g} \mathbf{c}$ where $\mathbf{c}$ is a vector of appearance parameters that controls both shape and texture. An AAM fitting on a target image is a nonlinear optimization problem where we seek to $arg_{min_{\mathbf{c}}} |\mathbf{I}_{image} - \mathbf{I}_{model}|^2$ updating $\mathbf{c}$ and pose. The training stage consists on estimate a Jacobian matrix that relates the texture difference and the corresponded additive updates that should be made to the model parameters $\mathbf{c}$. Figure 1 shows an AAM search where the initial estimate for the model position is given by AdaBoost [6] method.

3. Facial Expression Recognition

The facial expression recognition procedure is performed by firstly describing a set of faces using the AAM model [3] then projecting each appearance vector into hyperspace that maximize class separability using a multiclass



(a) 1^{st} (b) 2^{nd} (c) 3^{rd} (d) 6^{th} (e) Final (f) Orig.

Figure 1. Model Fitting.

Support Vector Machines (SVM) [1].

3.1. Support Vector Machines (SVM)

SVMs are maximal margin hyperplane classification methods that relies on results from statistical learning theory to guarantee high generalization performance. SVM is capable to solve linear and nonlinear classification problems. In the nonlinear case a kernel function is used to map the input data into the feature space, normally with higher dimensionality. The propose of SVM is to map this nonlinear data into a higher dimensional space, and make them linearly separable in that space. In this work, the facial expression SVM classification was achieved using a multi-class one-against-all voting scheme with a gaussian Radial Basis Function (RBF) kernel, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2}$, where the features data is normalized by mapping-it into the unitary hypersphere. The kernel parameters, i.e., γ and the misclassification penalty C, were found by cross-validation ranging the values in the interval $[1^{-12}, \ldots, 1^{12}]$.

3.2. Experimental Results

For the purpose of this work, an expression database was built, consisting in 21 individuals presenting 7 different facial expressions each, namely neutral expression, happiness, sadness, surprise, anger, fear and disgust. The dataset is therefore formed by a total of 147 colour images. The shape model was built using 58 annotated landmarks and the texture model was generated sampling around 47000 pixels using colour information. The dimensionality of appearance vector, **c**, changes as function of the retained variance on the three PCA of the AAM building process. 17, 29, 42, 70, 97, 133 are the dimension values, for 95%, 97%, 98%, 99%, 99.5% and 99.9% respectively of held variance in the training set used. Since, more dimensional data, not always produces better classification results, several experiences for each variance held dataset were performed. A leave-one-out cross-validation scheme was used to evaluate the classification. Table 1-top shows results of the confusion matrix using a multiclass SVM. The better classification results come from 95% of held variance. From the results, is shown that there is an effect of confusion between neutral/sad expressions and also between anger/disgust, suggesting appearance correlation between these two pairs of expressions. In fact, these results are not surprising, there are neuroscience studies [4] that support this idea. Eliminating the confusion effect, i.e. excluding the sadness and anger emotions, the better classification results, table 1-bottom, were achieved from the 99.5% held variance.

Table 1. Confusion matrices.

	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**

Overall recognition rate = 57.14%

	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**

Overall recognition rate = 80%

4. Conclusions

Standart AAM model formulation was used to describe face characteristics in a compact way. Several facial emotions were classified projecting the AAM appearance parameters to the hyperplane that maximizes class separability using a multiclass one-against-all SVM. With all the seven expressions it was achieved an overall recognition rate of 53.57%, but since the results emphasis appearance correlations between the pairs of expressions neutral/sad and anger/disgust, comproved by psico-physics studies, these correlated expressions were removed. Classifying five expressions shows that the recognition rate increases to about 80%.

References

[1] *The Nature of Statistical Learning Theory*. Springer-Verlag, N.Y., 1995.

[2] *Handbook of Cognition and Emotion*. John Wiley & Sons Ltd, 1999.

[3] M. D. Bouchra Abboud, Frank Davoine. Facial expression recognition and synthesis based on an appearance model. *Signal Processing Image Communication*, 2004.

[4] W. D. Killgore and D. A. Yurgelun-Todd. Activation of the amygdala and anterior cingulate during nonconscious processing of sad versus happy faces. *NeuraImage*, 2003.

[5] G. E. T.F.Cootes and C.J.Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[6] P. Viola and M. Jones. Rapid object detection using a boosted cascate of simple features. In *Computer Vision and Pattern Recognition*.