# Classification of Facial Expressions from Static Images

W. A. FELLENZ, J. G. TAYLOR, N. TSAPATSOULIS and S. KOLLIAS

Department of Mathematics, University of London, King's College, Strand, WCR2, United Kingdom
Computer Science Division, National Technical University of Athens, Zographou 15773, Greece

*Abstract:* We compare the performance and generalization capabilities of different low-dimensional representations for facial emotion classification from static face images showing happy, angry, sad, and neutral expressions. Three general strategies are compared: The first approach uses the average face for each class as a generic template and classifies the individual facial expressions according to the best match of each template. The second strategy uses a multi-layered perceptron trained with the backpropagation of error algorithm on a subset of all facial expressions and subsequently tested on unseen face images. The third approach introduces a preprocessing step prior to the learning of an internal representation by the perceptron. The feature extraction stage computes the oriented response to six odd-symmetric and six even-symmetric Gabor-filters at each pixel position in the image.

The template-based approach reached up to 75% correct classification, which corresponds to the correct recognition of three out of four expressions. However, the generalization performance only reached about 50%. The multi-layered perceptron trained on the raw face images almost always reached a classification performance of 100% on the test-set, but the generalization performance on new images varied from 40% to 80% correct recognition, depending on the choice of the test images. The introduction of the preprocessing stage was not able to improve the generalization performance but slowed down the learning by a factor of ten.

We conclude, that a template-based approach for emotion classification from static images has only very limited recognition and generalization capabilities. This poor performance can be attributed to the smoothing of facial detail caused by small misalignments of the faces and the large inter-personal differences of facial expressions exposed in the data set. Although the nonlinear extraction of appropriate key features from facial expressions by the multi-layered perceptron is able to maximize classification performance, the generalization performance usually reaches only 60%.

*Key-Words:* - facial analysis, emotion recognition, static face images, MLP CSCC'99 Proc.Pages:5331-5336

## 1 Introduction

During the last years, numerous architectures and algorithms for face recognition and expression recognition from static facial images have been proposed. Surveys of this field can be found in [1]. A general distinction into feature- and template-based approaches has been described by Brunelli and Poggio [2], but psychological experiments indicate that the human visual system processes faces at least to some extend holistically [3], favouring template-based approaches over feature-based techniques for their biological validity. We will therefore focus on advanced template-based techniques which can be further subdivided into supervised approaches (Average, MLP) and unsupervised techniques (PCA). The next section examines the use of the average template generated from the face images by summing up and normalizing the individual faces from each emotion class. Section three explores facial templates generated by a principal component analysis of the data-set. In section four a multi-layer perceptron trained with the backpropagation of error learning procedure is used to classify the facial images into the corresponding emotional expression classes. This approach is extended in section five by introducing a preprocessing stage which extracts oriented quadratur Gabor-wavelets at each pixel position in the image. The paper concludes with a discussion of the observed results.

## 2 The Averaged Templates

To investigate the recognition and generalization performance of the various methods several computer experiments were undertaken. The Image set, which was obtained from CMU, contains pictures of 20 different males and females. There are

32 different images (maximum size120x128) for each person showing happy, sad, neutral, and angry expressions, and looking straight to the camera, left, right, or up. The images with the highest resolution and straight facial orientation were normalized and cropped by a multi-scale head search, resulting in 77 face images of size 35x37. Four persons were excluded from further analysis due to a missing expression in the data-set or a failure of the normalization procedure to extract the head at the appropriate scale. The images in each emotional expression class were summed up and normalized to produce a generic template. Figure 1. Shows the four extracted templates which correspond to neutral, angry, happy, and sad expressions, respectively.

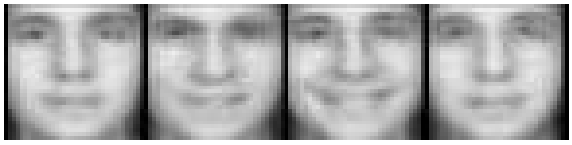

**Fig. 1** Emotion templates generated by averaging over all 16 different face images of the data-set. From left to right: neutral, angry, happy, and sad.

Although the templates are well defined and the exposed expression can be recognized by humans, the classification performance on the training-set was only 75%, which corresponds to three correct classifications out of four. In a further study the generalization performance to unseen images was examined, using a cross-validation procedure where one person was left out during the averaging and subsequently testing the left-out images on the emotion templates. The generalization performance only reached about 50% correct classifications, indicating the limited use of the averaged templates for general expression classification. A closer examination of the classification errors revealed small misalignments of the face images caused by head tilt during the exposure of the expressions, which was not compensated in the normalization procedure, and the individual peculiarity in emotion expression.

# 3 Principle Component Analysis

The use of Principle Component Analysis (PCA) for face recognition has been described by several research groups [3-9] in the last decade. The general idea behind this global approach is to extract the main information in the training set as represented by some template images that capture most of the variability in the data. This is achieved by projecting the input images onto orthogonal basis images, which have the property of allowing the best possible reconstruction of the training images. The general approach of PCA is to calculate the eigenvectors of the covariance matrix of the input data, and to use only the eigenvectors with eigen-values above some threshold to represent the input data and their principal directions. Since the calculation of the eigenvalues of a covariance matrix with large dimensional vectors is computationally intensive, this method can not in general be applied to image processing applications, where the size of the vectors equals the number of pixels of the training images.

## 3.1 PCA for Expression Classification

To allow for the fast computation of the principal components of face images, a different approach has been proposed: the principle components are not calculated from the raw image intensity, but from the covariance matrix of the input images themselves. Since this quadratic matrix is only of size N x N, where N is the number of training images, the fast computation of the K eigenvectors with the largest eigenvalues is possible. The extracted eigenvectors represent the combinations of the N input images which capture most of the variability in these images, and allow the reconstruction of all training images with the least mean squared error. The PCA method, which has been termed the Eigenface method due to the similarity of the appearance of the template images to ‘Ghost’-Faces, relies on the assumption that a low-dimensional representation of the face images using a small value of K, much smaller than N, suffices to capture most variation in the training images. This is not true in general, since the training images could be very dissimilar, resulting in a poor representation if only the eigenfaces with the K largest eigenvalues are considered. Therefore, it is necessary to align the images to a general viewpoint prior to the Eigenface decomposition by translating, scaling, and rotating the faces to a reference position. It has been demonstrated [6] that although choosing the K largest eigenvalues is optimal for identifying physical categories of faces like sex, it is not optimal for recognising faces. Instead eigenvectors with smaller eigenvalues may provide a better representation for recognition. The study by Turk and Pentland [4] shows that a small number of distinct Eigenimages suffice to recognise all training images and slightly different test images which vary in illumination and pose.

**Fig. 2** The largest 36 principal components generated from the data-set of 64 face images showing neutral, angry, happy and sad facial expressions.

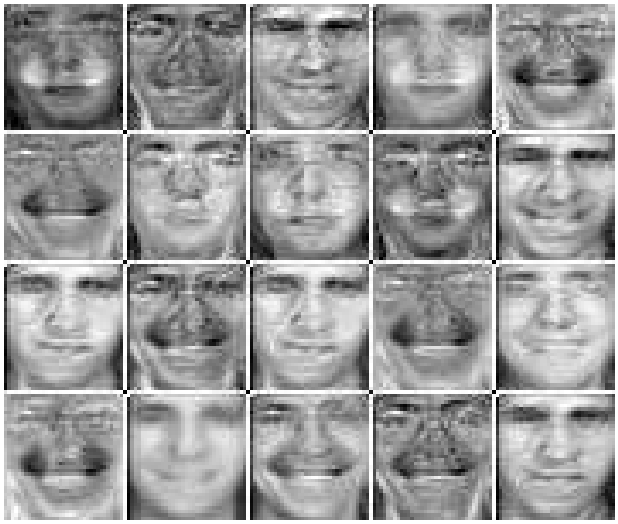

**Fig. 3** Weights of 20 hidden layer neurons of a MLP-network of size (1295x20x4) for recognising the facial expressions neutral, angry, happy, and sad.

We adapted the general eigenface-decomposition procedure to produce a low-dimensional representation of all faces in the data-set (Figure 4). As can be seen from the extracted principal components, most of the variation in the training-data is captured in the eigenfaces with the 12 to 18 largest eigenvalues. The recognition performance of the individual faces and the exposed expressions using only the 10 largest principal components was almost perfect, indicating the superior performance on a face recognition task. However, since the representation is well tuned to the individual details of the images from the training-set, allowing the best possible reconstruction of the original images from the principal components, it is not able to generalize well to unseen faces and their expressions. Therefore this scheme has to be modified to incorporate a better generalization performance. We conclude that the use of the low-dimensional PCA representation for classifying of facial expressions is more suited for recognizing and reconstructing known facial expressions than to generalize to unseen faces.

# 4 The Multi-Layer Perceptron

The use of supervised learning techniques employing a multi-layer perceptron (MLP) for face recognition and face perception has been adopted in many systems [10, 11]. The general idea is to use a feedforward neural network with one or more intermediate layers which are fully connected to an output layer, where each output neuron represents one predefined target output, and the system is allowed to selforganize the appropriate weights between input to hidden layer, and hidden to output layer by minimising the error at the clamped output units. This is usually achieved by the powerful backpropagation of error algorithm, which gradually decreases the overall error for all known input to output combinations by adjusting the intermediate weights of the network. The output units can represent each individual for a face recognition task, or physical categories like female and male in a gender recognition task [10]. We modified a MLP network with four output units and a hidden layer to extract the facial expressions from the data-set. This image-set was split into a training set containing nine images, and two further sets for validation and testing of size five each.

## 4.1 Classification Performance of the MLP

After about 850 learning cycles using the backpropagation of error algorithm the network with 35x37 input units, 20 hidden units, and four output units converged, and was able to recognise all expressions from the training set with 100% accuracy. The generalization performance was tested with 5 unseen images from the test set, and reached up to 78% correct classification of the exposed expressions. Since 25% correct is chance level, the network can classify three out of four, which is a remarkable performance considering the variation in interpersonal emotion expression and the intra-personal similarity of some of the face images. However, if the test set was chosen at random in a cross-validation study, lower levels of

generalization performance were measured (about 40-60% correct classification). Depicted in figure 3 are the 20 images of the learned weights of the hidden neurons of the (1295, 20, 4) - MLP network. Some of the images have been inverted to reveal a more realistic impression of the extracted features. Compared to the principal components depicted in figure 2, the hidden layer representation of the MLP is less tuned to the individual facial details of the training-set, but is more related to the emotional content of the face images. This is apparent from the light and dark shades around the wrinkles of the mouth, indicating the importance of this feature for emotion classification (e.g. the first, fourth, and last image in the upper row). Another apparent feature is the position of the eyebrows, which indicates an angry facial expression (e.g. the third image in the upper row and the last image).

### 4.2 Compression of the Representation

In a second study we reduced the hidden layer representation to five neurons to reveal the most critical features needed for facial emotion classification from static images. Again the MLP-network converged after 800 iterations, reaching 100% correct classification performance on the images from the training set. The compression ratio for the reduced hidden layer representation is 64/5 = 12.8, since all images from the training-set can be correctly classified. The generalization performance was comparable to the previous network with 20 hidden neurons, but the hidden layer representation depicted in figure 4 shows a more defined feature set. The third and fourth neurons show similarity to an ‘eyebrow’-detector, which is an important feature for face expression recognition. Closer inspection of the position of both eyebrows show a small displacement upwards for the third and downwards for the forth neuron compared to the average face. Both displacements correspond to happy and angry expressions, respectively, which is apparent from the distribution of the neuron‘s weights. The first and the last neuron are selective for regions of the mouth and seem to measure the curvature of the lips. This feature is present in most of the images of hidden layer neurons trained in the expression recognition task, suggesting its general importance for face expression perception. The rotation visible in the second image is caused by the rotation of some of the training faces and displays the perturbation of the network weights by an artifact.

**Fig. 4** Weights of five hidden layer neurons of a MLP-network of size (1295x5x4) for recognizing the facial expressions neutral, angry, happy, and sad.

## 5 Preprocessing by Gabor-Wavelets

To improve the generalization performance of the MLP-network we introduced a preprocessing stage, which consists of filtering the face images with a set of oriented quadrature phase Gabor-wavelets [12,13]. The response of the 12 oriented Gabor-wavelets to a neutral face from the data-set is depicted in figure 5, and the response to a happy facial expression of the same person is shown in figure 6. As can be seen from the orientation maps, most facial expression information is contained in the horizontal filter responses of the mouth and the eyebrows, although some important information may as well be found in the adjacent feature-maps. For example, the upward movement of the mouth during a smile can easily be detected in the orientation maps next to the horizontal map. However, the vertical orientation does not contribute as much as the horizontal ones and could be left out to speed up the learning procedure.

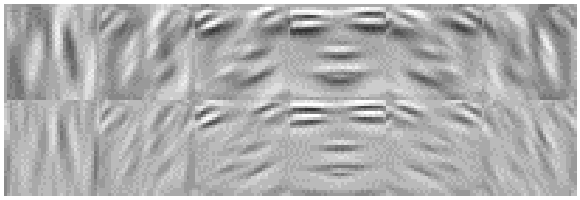

**Fig. 5** Preprocessed face image exposing a neutral expression by filtering the image with six oriented odd-symmetric Gabor wavelets (upper row) and six even-symmetric wavelets (lower row). White shades correspond to a positive filter response, black to a negative response, gray corresponds to zero level.

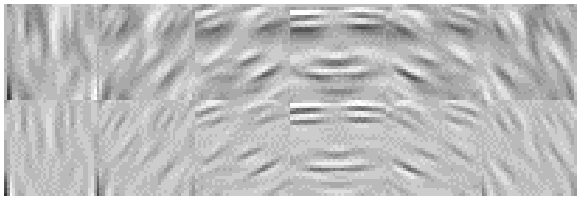

**Fig. 6** Preprocessed face exposing a happy expression, (see Fig. 5 for details).

The training procedure for the MLP-network was the same as before, but now the 12 feature maps for the face images were used as the training input to the network. As before, the converged network was able to correctly classify all images from the training-set. A sample image of a hidden neuron is depicted in figure 7, showing the adoption of the Gabor-wavelet representation of the input by the hidden layer neurons. However, no improvements on the generalization performance to novel images were observed.

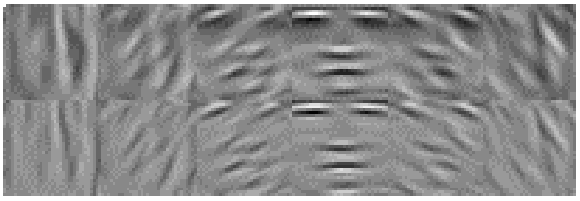

**Fig. 7** Weights of a single hidden layer neuron of an MLP-network of size (14700x5x4) using the Gabor-filtered face images as training patterns.

## 6 Conclusion

In the previous sections we explored and compared the classification and generalization performance of various supervised and unsupervised techniques for emotion classification from static images. The template-based approach for emotion classification from static images has only very limited recognition and generalization capabilities. Its poor performance may be caused by the smoothing of important individual facial detail, e.g. the curling of the forehead, and by small misalignment of the faces. Furthermore, the expressions exposed in the data set showed large inter-personal differences, and could sometimes not be recognized by humans. The PCA-representation showed excellent classification and reconstruction performance on the training-set but does not allow generalizing to novel faces. The MLP-network trained with backpropagation on the other hand showed perfect classification and an acceptable generalization performance, if the test images were not too dissimilar from the training faces. We conclude, that the nonlinear extraction of appropriate key features from facial expressions by the multi-layer perceptron is able to maximize classification performance on almost any data set. The generalization performance to novel images critically depends on a good alignment of the facial images, an expressive facial emotion display of the individuals, and a sufficient resolution of the face images, allowing subtle detail to be extracted. In summary the limited generalization performance in the present study was caused by three limitations:

1. Limitations of the data set:
   a) low-resolution faces of size 35x37 pixel
   a) only three emotions (sad, angry and happy)
   b) exposed expressions are not recognizable
2. Limitation of the head normalization procedure
3. Restriction to static images

To improve the generalization performance on novel faces, several modifications can be introduced. To improve the alignment of the static images, a second normalization stage can be used which extracts key-points of the face like the corners of the mouth and both eyes and realigns the faces to a standard position. A recent study has shown [14] that using Gabor-wavelet coefficients at a limited number of fiducial points, which were selected by hand, can enhance generalization performance to 90% compared to 70% if only the geometric positions were used. A computational expensive way to improve generalization performance is by using the temporal evolution of the facial expression to extract an optical flow field [15,16] or a temporal difference image [17]. A further approach uses the detailed geometry of facial muscle activation [18, 19] and reconstructs the emotional expression employing a codebook of all possible combinations. However, no single technique will reach sufficient generalization performance considered separately. Therefore, it is appropriate to use all information sources that are available to improve the stability and performance of the system by incorporating static images, image sequences and speech information [20].

*References:*


[1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings IEEE*, Vol. 83, No. 5, 1995, pp. 705-740

[2] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, 1993.

[3] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 1990, pp. 103-108.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, 1991, pp.71-86.

[5] G.W. Cottrell and J. Metcalfe, "EMPATH: Face, emotion, and gender recognition using holons," In R.P. Lippman, J. Moody and D.S. Touretzky

(Eds.) *Advances in Neural Information Processing Systems 3*, San Mateo: Morgan & Kaufman, 1991, pp. 564-571

[6] A.J. O'Toole, H. Abdi, K.A. Deffenbacher and D. Valentin, "A low-dimensional represen-tation of faces in the higher dimensions of the space," *Journal of the Optical Society of America A*, Vol. 10, 1993, 405-411.

[7] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *In Proceedings of the European Conference on Computer Vision*, 1996

[8] P.J.B. Hancock, V. Bruce and A.M. Burton, "Testing principle component representations for faces," *Proc. 4th Neural Computation and Psychology Workshop*, London, 9-11 April 1997, Eds. J.A. Bullinaria, D.W. Glasspool, and G. Houghton, pp. 84-97, London, Springer, 1997

[9] B. Moghaddam, W. Wahid and A. Pentlant, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," *Proc. 3nd IEEE Intl Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, 1998

[10] M.S. Gray, D.T. Lawrence, B.A. Golomb and T.J. Sejnowski, "A perceptron reveals the face of sex," *Neural Computation*, Vol. 7, No. 6, 1995, pp. 1160-1164.

[11] N. Intrator, D. Reisfeld, and Y. Yeshrun, "Face Recognition using a Hybrid Supervised/ Unsupervised Neural Network," *Pattern Recognition Letters*, Vol. 17, 1996, pp. 67-76

[12] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. 36, No. 7, 1988, pp. 1169-1179

[13] J. Buhmann, J. Lange and C. von der Malsburg, "Distortion invariant object recognition by matching hierarchically labeled graphs," *In IJCNN International Conference on Neural Networks*, Washington, DC, Vol. 1, pp. 155-159, 1989

[14] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptrons," *In Proceedings of the 3rd IEEE International Conference on Automatic Face & Gesture Recognition*, Nara, Japan, 1998

[15] I. Essa, T. Darrell and A. Pentland, "Tracking facial motion," *Proceedings of the Workshop on Motion of Nonrigid and Articulated Objects*, IEEE Computer Society, pp. 36-42, 1994.

[16] Y. Yacoob and L..S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 6, pp. 636-642, 1996.

[17] C. Padgett and G.W. Cottrell, and R. Adolphs, "Categorial perception in facial emotion classification," *manuscript*, 1998

[18] M. Stewart Bartlett, P.A. Viola, T.J. Sejnowski, B.A. Golomb, J. Larsen, J.C. Hager, P. Ekman, "Classifying facial action," *Advances in Neural Information Processing Systems 8*, D. Touretzky, M. Mozer, and M. Hasselmo (Eds.), MIT Press, Cambridge, MA, 1996.

[19] J.J.-J. Lien, T. Kanade, J.F. Cohn and C.C. Li, "A multi-method approach for discriminating betweem similar facial expressions, including expression intensity estimation," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1998, pp. 853-859

[20] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. Taylor, "Emotion recognition in human-computer interaction," *submitted for publication to the IEEE Signal Processing Magazine*, January 1999.