Artificial Neural Networks: Biological Inspirations - ICANN 2005 (ICANN), Warsaw, LNCS 3696, vol. I, pp. 569-574, Springer 2005

Classification of Face Images for Gender, Age, Facial Expression, and Identity[1]

Torsten Wilhelm, Hans-Joachim Böhme, and Horst-Michael Gross

Department of Neuroinformatics and Cognitive Robotics
Ilmenau Technical University, P.O.Box 100565, 98684 Ilmenau, Germany

Abstract. In this paper we compare two models for extracting features from face images and several neural classifiers for their applicability to classify gender, age, facial expression, and identity. These models are i) a description of face images by their projection on independent base images and ii) an Active Appearance Model which describes the shape and grey value variations of the face images. The extracted feature vectors are classified with Nearest Neighbor, MLP, RBF and LVQ networks, and classification results are compared.

1 Introduction

A growing number of applications rely on the ability to extract information about people from images. Examples are person identification for surveillance or access control, the estimation of gender and age for building user models or facial expressions recognition, which can give valuable information for the evaluation of man-machine interfaces. As the mentioned recognition tasks have been addressed in isolation in the past, there often exists a variety of methods for each. The presented work was done in the context of building a man-machine interface for a mobile service robot [5], where all the above mentioned information is of great interest. Furthermore, our hope was to identify one method that could be used universally.

2 State of the Art

A commonly used method for face image analysis is the subspace projection of the image data, where the subspace can be spanned by principal components, independent components of the training data. This method was used for a vast amount of approaches for person identification and automatic facial expression analysis [1]. Another widespread method for person identification and for gender estimation is the Elastic Graph Matching technique [10] [7]. Elastic Graph Matching describes faces in terms of spatial frequencies in local image areas, where the relation between these areas is defined by a graph structure. These models are adaptive and can adjust themselves to some degree to variations in the image data. Active Appearance Models have been used for person identification and facial expression analysis [2] [3]. They describe the statistical variations of shape and grey values in the training data and adapt themselves in an iterative process to a given face image. Up to now, very little work was reported on age estimation from image data.

[1] This work is partially supported by TMWFK-Grant # B509-03007 to H.-M. Gross

3 Dataset coding age, gender, identity, and facial expressions

There exists a variety of face databases designed for single classification tasks, e.g. the Cohn Kanade database for facial expressions [8]. However, we wanted to test the performance of our methods for classification tasks as diverse as gender, age, facial expressions, and identity. To eliminate the influence of varying quality of the image data, we recorded our own database according to our requirements. This database consists of two parts. The first part used for the classification of age, gender, and identity contains 70 people with 7 images each, with neutral facial expression, different illuminations, and small deviations in head orientation. Identities are equally distributed in the age range between 10 and 60 and equally distributed over genders. The second part consists of 30 identities with 7 images each, which represent the basic emotions happiness, sadness, surprise, fear, anger, disgust and neutral as identified by Ekman in [4], see Fig. 1. Since evoking facial expressions with movies was shown to produce mixtures of basic emotions [6], we decided to ask probands to pose facial expressions according to the seven basic emotions. As people's ability to pose facial expressions varies significantly, all the recorded images were manually classified by 10 people and only a subset of all images was used where at least 7 people agreed on the facial expression. This problem could be avoided, if the image data was labeled with FACS codes, describing the activity of facial muscles instead of basic emotions. However, since FACS coding is very time consuming and has to be done by trained personnel, it was not yet possible to obtain FACS codes for our data.

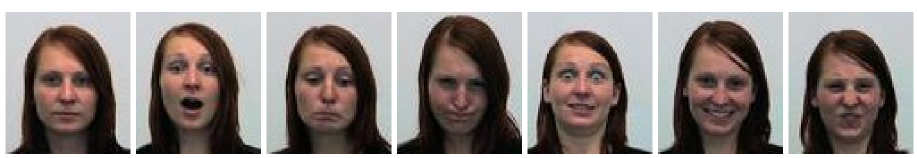


Fig. 1. Examples of the used data set for facial expressions. From left to right: neutral, surprise, sadness, anger, fear, happiness, disgust.

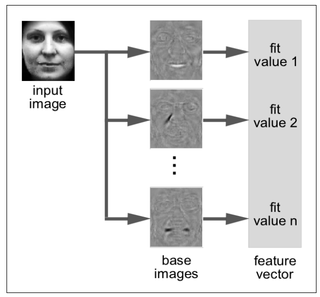
4 Feature Extraction

To provide the feature extraction methods with normalized data, we manually labeled the position of facial landmarks. We analyzed Independent Component Analysis (ICA) and Active Appearance Models (AAM) for feature extraction.

4.1 Independent Component Analysis

For ICA, only the centers of the eyes are used as facial landmarks. The ICA model depends on highly accurate aligned image data as the model does not adapt to a given face image. Thus, we applied affine transformations such that the center of the eyes are on the same position in every image. The size of these normalized images is 60×70 pixels. For building the ICA model, an observation matrix was built by using the vectorized images as rows. On this matrix we applied ICA and obtained independent base images for the data, which represent local facial features, see Fig. 2(a). Any given normalized face image is represented as linear combination of independent base images, and the fit values constitute the data to be classified. For a more detailed description see [9]. The data flow when using the ICA model is depicted in Fig. 2(b).

(a)



(b)

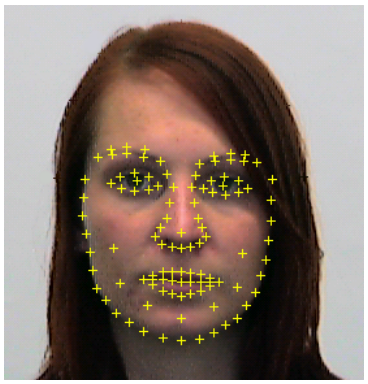
Fig. 2. (a) Example of an independent base image. (b) When using the ICA model, a normalized input image is projected on the independent base images. The fit values for the base images form the feature vector to be classified.

4.2 Active Appearance Models

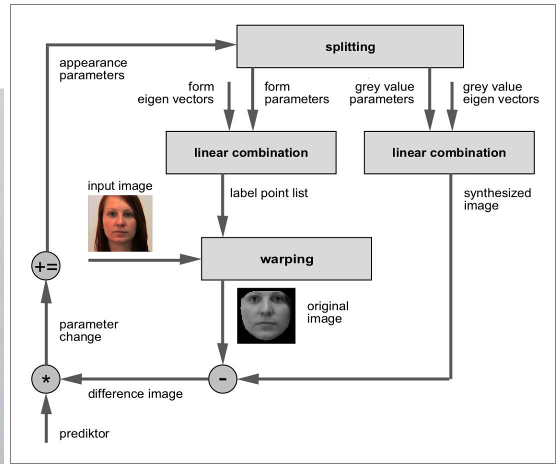
The facial landmarks for the Active Appearance Models (AAM) consist of 116 points along dominant outlines in the face, see Fig. 3(a). To construct an AAM, the mean shape of the training data is computed and the shape variation is determined by principal component analysis. In the next step, the training images are warped to the mean shape. In the same way as with the shape model, the mean grey value face is computed and the grey value variation is determined by principal component analysis. Finally, a predictor matrix is estimated by varying single appearance parameters and averaging their effects on the difference image. For details on Active Appearance Models see [2]. The data flow when using the AAM is depicted in Fig. 3(b). After adaptation of the model to a given face image, the resulting appearance parameters describe the shape and the grey value distribution of the given face and are used as feature vector for classification.

5 Classification

We used a leave-n-out strategy to train and test the classifiers. The partitioning was such, that every identity was in the test dataset exactly once. For gender, the test dataset consisted of 2 identities (male and female), for age of one identity per age group and for facial expressions of one person showing all facial expressions. For person identification we used 3 images from each person for training, 2 for validation, and 2 for testing. The results were averaged over the multiple training cycles performed for each recognition task. We compared the following network types: Multi Layer Perceptron (MLP), Radial Basis Function (RBF), Nearest Neighbor (NN), and Generalized Learning Vector Quantization (GLVQ). The number of inputs corresponds to the number of extracted features, that is, ICA fit values or appearance parameters, respectively. The MLPs had two trainable layers with 40 neurons in the hidden layer. The GLVQ network used 20 neurons per class. The centers of RBF networks were initialized by GLVQ training. GLVQ and RBF networks operated on normalized, NN and MLP on unnormalized feature vectors.



(a)



(b)

Fig. 3. (a) Landmarks used for the construction of the Active Appearance Model. (b) Usage of the Active Appearance Model. The iterative search process begins with appearance parameter vector **0**, that is, with mean shape and mean grey value distribution. The position is initialized by the center of the eyes as with the ICA model. The input image is warped to the current shape to produce the form normalized *original image*. On the other hand, the current grey value parameters are used to produce the form normalized *synthesized image*. From the difference of these images, a parameter change is estimated for each appearance parameter with the goal to minimize the energy of the difference image. The search process converges, when this parameter change is **0**.

6 Results and Conclusions

Recognition rates are shown in Fig. 4. For gender and facial expressions, recognition rates are promising. Here, the best results were obtained with AAMs and MLP classifiers or ICA with Nearest Neighbor classifiers, respectively. For both feature extraction methods (ICA and AAM), the results for age classification are only slightly better than guessing (20% due to the used 5 classes). From the confusion matrixes it can be seen, that it is possible to distinguish young and old people, Table 1. For person identification it is often not feasable to have a fixed gallery represented by a neural classifier. Alternatively, two models either ICA fit values or active appearance parameters can be compared by the normalized dot product. Therefore, the false acceptance and false rejection rates were recorded for different similarity thresholds, see Fig. 5. Both methods achieve approximately the same equal error rates. Although the best recognition rates for gender and age estimation were obtained with AAMs, in the final system we deploy ICA with Nearest Neighbor classifiers only. This is a trade-off between recognition rates and processing time needed to adapt the AAM to the input image, which is about 2800ms compared to about 20ms for the subspace projection with ICA. Since the AAM proved its capability for various recognition tasks, our future work will focus on optimizing the AMM for processing speed. In contrast to the ICA subspace projection, which relies on acurately aligned frontal views, the AAM is also able to handle and estimate head pose, provided this image variation is present in the training data.

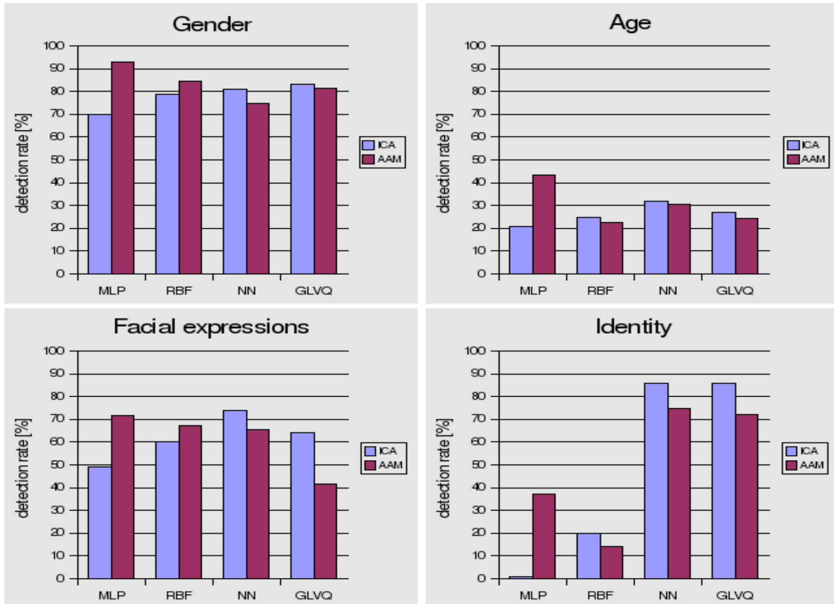


Fig. 4. Recognition rates for gender, age, facial expressions and identity on validation data for feature extraction with ICA or AAM, respectively. The best recognition rates w.r.t. gender, age and facial expression were obtained with AAM and MLP or ICA and NN, respectively. The high recognition rates for AAM with MLP classifiers suggest, that the AAM produces appearance parameters which are in contrast to the ICA fit values well clustered according to the recognition tasks. For person identification the ICA feature extraction performs significantly better than the AAM. Here MLP and RBF networks fail, because of the large number of clusters. However, in the final system, person identification is done by comparing two models and using a similarity threshold for acceptance or rejection, see Fig. 5.

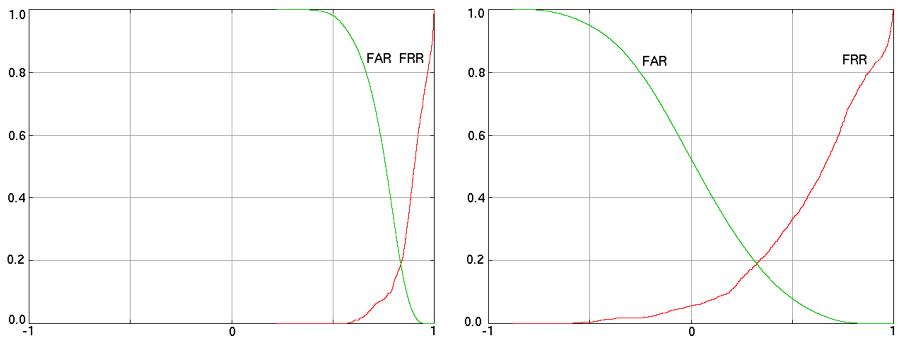
	10	**20**	**30**	**40**	**50**
10	48	15	18	10	7
20	9	36	22	6	25
30	18	32	20	15	13
40	6	19	29	23	21
50	7	32	21	19	19

(a)

	10	**20**	**30**	**40**	**50**
10	55	11	13	16	3
20	9	40	29	8	12
30	20	19	23	19	17
40	13	7	23	33	22
50	4	10	15	22	47

(b)

Table 1. Confusion matrixes for age estimation. Horizontal: true class, vertical: estimated class (a) ICA + NN (b) AAM + MLP. Each age intervall contained 98 images to be classified. It can be seen, that both methods are roughly able to distinguish young from old people. Thus, when using the system only two or three clusters should be used.



(a) (b)

Fig. 5. False acceptance (FAR) and false rejection rate (FRR) curves for person identification with (a) ICA (b) AAM. The equal error rates are approximately the same with 0.2 for ICA and 0.19 for AAM, which suggests that both methods are equally well suited for person identification.

References

1. M.S. Bartlett. *Face image analysis by unsupervised learning*. Kluwer Academic Publishers, 2001.
2. Cootes, T.F., Edwards, G.J., and Taylor, C.J. Active appearance models. *Lecture Notes in Computer Science*, 1407:484–??, 1998.
3. Edwards, G.J., Taylor, C.J., and Cootes, T.F. Learning to identify and track faces in image sequences. In *ICCV*, pages 317–322, 1998.
4. Ekman, P. and Friesen, W.V. *Unmasking the face. A guide to recognizing emotions from facial clues*. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
5. Gross, H.-M., Boehme, H.-J., Key, J, and Wilhelm, T. The perses project - a vision-based interactive mobile shopping assistant. *Künstliche Intelligenz*, 4:34–36, 2000.
6. Gross, J.J. and Levenson, R.W. Emotion elicitation using films. *Cognition and Emotion*, 9:87–108, 1995.
7. Hong, H., Neven, H., and v. d. Malsburg, C. Online facial expression recognition based on personalized gallery. In *Intl. Conference on Automatic Face and Gesture Recognition*, pages 354–359. IEEE Comp. Soc., 1998.
8. Kanade, T., Cohn, J.F., and Tian, Y. Comprehensive database for facial expression analysis. *In Proc. of the Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG'00), Grenoble*, pages 46–53, 2000.
9. Wilhelm, T. and Backhaus, A. Statistical and neural methods for vision-based analysis of facial expressions and gender. In *In Proc. IEEE Int. Conf. on System Man and Cybernetics*, pages 2203–2208, 2004.
10. Wiskott, L., Fellous, J.-M., Krüger, N., and von der Malsburg, C. Face recognition and gender determination. In *Int. Workshop on Automatic Face- and Gesture-Recognition, Zürich, June 26-28, 1995*, pages 92–97, 1995.