Components for Face Recognition

Bernd Heisele
Honda Research Institute USA, Inc.
Boston, USA
bheisele@honda-ri.com

Takamasa Koshizen
Honda Research Institute Japan, Co. Ltd.
Wako-shi, Japan
koshiz@jp.honda-ri.com

Abstract

We present a method for automatically learning a set of discriminatory facial components for face recognition. The algorithm performs an iterative growing of components starting with small initial components located around preselected points in the face. The direction of growing is determined by the gradient of the cross-validation error of the component classifiers. In experiments we analyze how the shape of the components and their discriminatory power changes across different individuals and views.

1. Introduction

In component-based face recognition the classification is based on local components as opposed to the global approach where the whole face pattern is fed into a single classifier. The main idea behind using components is to compensate for pose changes by allowing a flexible geometrical relation between the components in the classification stage. In addition, component-based recognition is more robust against local changes in the image pattern caused by partial occlusion or shadows. In [3] face recognition was performed by independently matching templates of the eyes, the nose and the mouth. A similar approach with an additional alignment stage was proposed in [1]. In [7] recognition was based on Gabor wavelet coefficients that were computed on the nodes of a 2D elastic graph. A comparison between global and component-based face recognition systems using Support Vector Machines (SVM) was published in [4].

The main difficulty in component-based object recognition is the selection of the components, i.e. how to find discriminatory components that allow to distinguish a particular object from other objects. In previous work [5] we introduced an algorithm that learns rectangular facial components for a face detection system. The algorithm starts with small initial components located around preselected points on the face. Each component is grown iteratively; the direction of growing is controlled by an SVM error bound of the component classifier. In the present paper we extend this method to the multi-class problem of face recognition and replace the error bound by the cross-validation error. Not only should the cross-validation error give us a better estimate of the prediction error it also makes the technique applicable to other types of classifiers besides SVMs. For every person in the training database we determine pose-specific sets of components. In our experiments we investigate how the learned components change across individuals and poses. This information might be relevant for a wide variety of face recognition and verification systems.

The outline of the paper is as follows: The image data is described in Section 2. In Section 3 we explain the method for learning components. Section 4 contains the experimental results and the discussion. Section 5 concludes the paper.

2. Face Data

Our method for learning components requires the extraction of corresponding components from a large number of training images. To automate the extraction process we used a set of textured 3D head models with known point-wise correspondences. The 3D head models were computed from image triplets (front, half profile, profile view) of six subjects using the morphable model approach described in [2]. Fig. 1 shows an original image and a rendered image of each of the six subjects in the database. Approximately 10,900 synthetic faces were generated at a resolution of 58×58 by rendering the six 3D face models under varying pose and illumination. The faces were rotated in depth from 0° to 44° in 2° increments. They were illuminated by ambient light and a single directional light pointing towards the center of the face. The directional light source was positioned between -90° and 90° in azimuth and between 0° and 75° in elevation. Its angular position was incremented by 15° in both directions. Example images with different pose and illumination settings are shown in Fig. 2.

To build the cross-validation set we rendered the 3D head

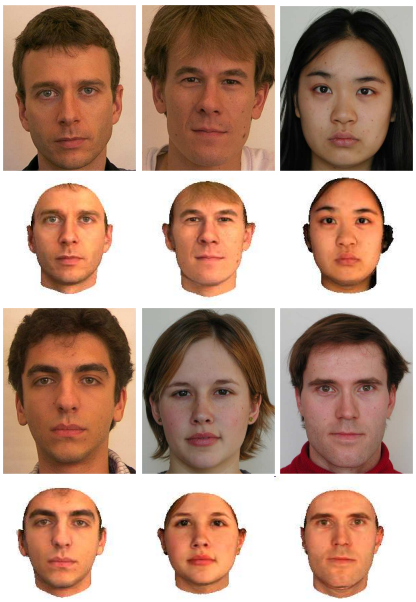

Figure 1. Original images and synthetic images for all six subjects in the database. The face images are arranged according to the ID numbers, starting with person 1 in the upper left corner and ending with person 6 in the lower right corner.

Figure 2. Examples of synthetic faces from the training set. Upper row shows the variations in pose and the bottom row shows variations in lighting.

models for slightly different viewpoints and illumination conditions. The faces were rotated in depth from 1° to 45° in 2° steps. The position of the directional light source varied between -112.5° and 97.5° in azimuth and between -22.5° and 67.5° in elevation. Its angular position was incremented by 30° in both directions. In addition, each face was tilted by $\pm 10^\circ$ and rotated in the image plane by $\pm 5^\circ$. The validation set included about 18,200 face images, some examples are shown in Fig. 3.

Figure 3. Examples of synthetic faces from the cross-validation set. The faces in the cross-validation set were tilted by $\pm 10^\circ$ and rotated in the image plane by $\pm 5^\circ$.

3. Learning Components

An intuitive choice of components for face recognition is the eyes, the nose and the mouth. However, it is not clear what exactly the size and shape of these components should be and whether there are other components which are equally important for recognition. Furthermore, we would like to quantify the discriminatory power of each component and analyze how the optimal set of components changes over pose and across different subjects. This can be accomplished by an algorithm for learning components which was developed in the context of face detection [5]. The algorithm starts with a small rectangular component located around a preselected point in the face (e.g. center of the left eye). The component is extracted from each face image to build a training set. A component classifier is trained according to the one-vs-all strategy, i.e. the components of one person are trained against the components of all other people in the database. We then estimate the prediction error of each component classifier by cross-validation. To do so, we extract the components from all images in the cross-validation set based on the known locations of the reference points. Analogous to the training data, the positive cross-validation set includes the components of one person and the negative set includes the components of all other people. After we determined the recognition rate on the cross-validation set (CV rate) of the current component classifier, we enlarge the component by expanding the rectangle by one pixel into one of four directions: up, down, left or right.

Again, we generate training data, train an SVM and determine the CV rate. We do this for expansions into all four directions and finally keep the expansion for which the CV rate increases the most. This process can be repeated for a preselected number of expansion steps. We ran our experiments on fourteen components, most of them located in the vicinity of the eyes, nose and mouth (see Fig. 4).

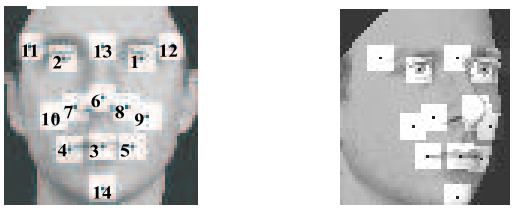

Figure 4. The initial fourteen components for a frontal and rotated face.

4. Experiments

The rotation in depth of the faces in the training and validation sets ranged from $0°$ to $44°$, with increments of $2°$. We split the data into three subsets: $[0°, 14°]$, $[16°, 30°]$ and $[32°, 44°]$, which in the following are referred to as pose intervals 1, 2, and 3, respectively. Each subset included about 630 training and 1060 validation images of each person. To speed-up computation, we randomly removed two thirds of the images in the cross-validation set, leaving around 350 images in each validation subset. For each person and each pose interval we trained a set of fourteen component classifiers, resulting in 252 component classifiers overall.

In a first experiment we determined the CV rate depending on the width and height of a symmetric component. Symmetric components are rectangular components which are centered on a reference point, i.e. their expansions to the left and right are identical, as are the expansions up and down. The dependency on only two variables allows a 3D visualization of the CV rate. We computed the CV rate for SVMs with second-degree polynomial kernel for all sizes of a symmetric component between 5×5 and 21×21 pixels. As features we used the histogram-equalized gray values of the component patterns. Some results are depicted in Fig. 5. The first four rows show the CV rates for the components 2 (right eye), 3 (center of the mouth), and 10 (right cheek) for pose interval 3 ($32°$ to $44°$). The last two rows show how the CV rate of component 3 changes across the pose. The surfaces are relatively smooth with few local maxima. It can be expected that gradient-based methods, such as the one described in the previous Section, can be successfully applied to find the maxima. The diagrams also show that the CV rate

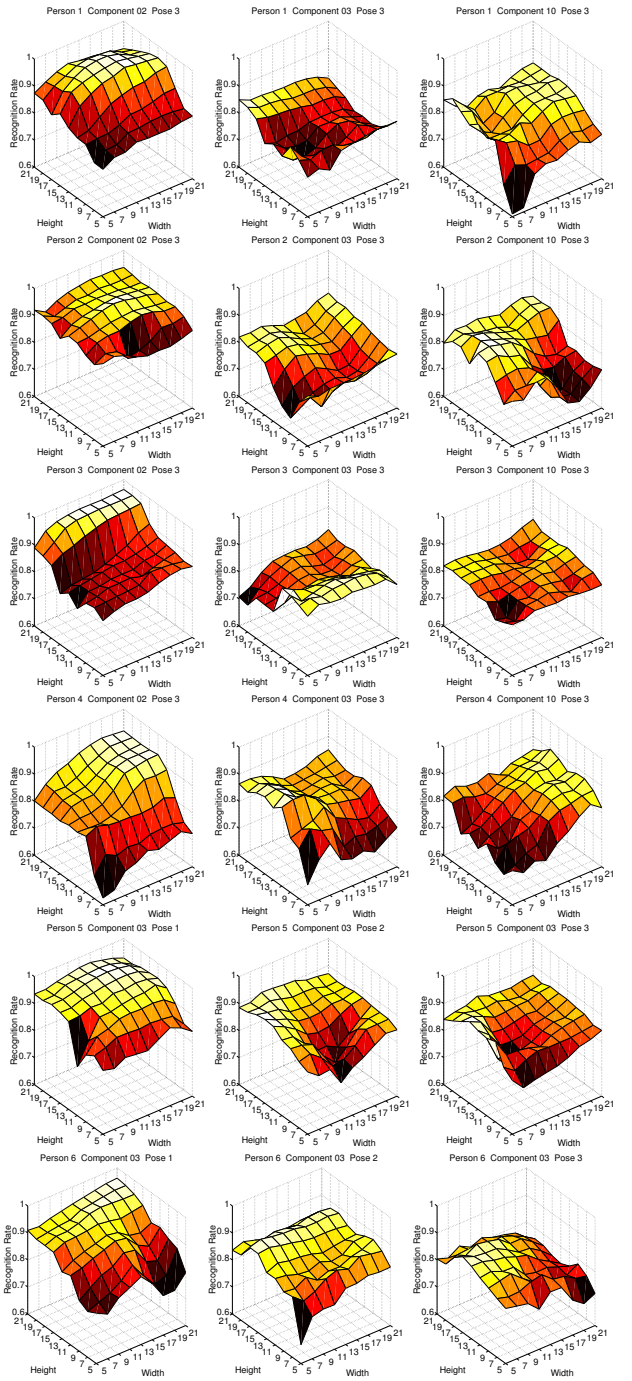

Figure 5. The CV rate for symmetric components. The first four rows show the CV rates for the components 2 (right eye), 3 (center of the mouth), and 10 (right cheek) for the pose interval $[32°, 44°]$. The last two rows show how the CV rate changes across the pose intervals $[0°, 14°]$, $[16°, 30°]$ and $[32°, 44°]$.

for a given component strongly changes between subjects, indicating that person-specific components yield better discrimination results than a universal set of components. Another important observation is that the results for a given component/subject combination vary over pose, which suggests the use of view-specific components.

In a second experiment we applied the algorithm described in Section 3 to learn a set of rectangular components. The initial size of our 14 components was set to 9×9 pixels. The number of iterations in the growing process was limited to 10, resulting in 10 components of different sizes and with different CV rates. Of the 10 components we selected the one with the maximum CV rate as our final choice. As for the previous experiment, we trained SVMs with a second-degree polynomial kernel on the histogram-equalized gray values of the components. Fig. 6 shows the learned components across different subjects and views. The average intensity of the component encodes the CV rate, bright values indicate a high CV rate. The pictures show that the CV rates decrease with increasing rotation. This effect is more prominent for components on the left side of the face—the side to which the faces were rotated—than for components on the right side. For the given system, the optimal pose interval for recognizing faces is the near frontal interval, which is similar to results reported in psychophysical experiments on face recognition in humans [6]. The component 6, located on the tip of the nose, and the components 7 and 8 around the nostrils are relatively weak. Most likely because the image pattern around the nose strongly varies with changes in illumination and pose. It can also be seen that the shapes of the components change across subjects (for fixed pose) and across pose (for fixed subject). The bar diagrams with the CV rates are shown in Fig. 7 and Fig. 8 arranged according to subjects and pose, respectively.

Table 1 shows the ranking of the components according to the average CV rate and the standard deviations of the CV rate across pose and across subject. The top five components according to the ranking by the CV rate are the outer right brow, both eyes and the chin. The left eye is the only component among the top eight components, which is located on the left half of the face. The discrepancy between the CV rates for components on the right and left side of the face shows clearly for the brow components (11, 12), which are ranked first and ninth. Slightly unexpected, the mouth components 3, 4 and 5 are only ranked in the middle, despite the complete absence of variations in facial expression in our training and validation sets. Interestingly, the nose component has the smallest σ across the pose and the largest σ across the subjects.

We will conclude the discussion with a some remarks on the applicability of the results. Learning and ultimately using view- and person-specific components for recogni-

Figure 6. The shapes of the learned components across pose intervals and across subjects. The brighter the component, the higher its CV rate. The subjects are arranged along the rows starting with person 1 in the first row. The columns represent the three pose intervals. From left to right: $[0°, 14°]$, $[16°, 30°]$ **and** $[32°, 44°]$**.**

Comp	CV Rate	Comp	σ_{Pose}	Comp	σ_{ID}
11	0.942	6	0.009	5	0.014
2	0.932	11	0.019	14	0.018
1	0.918	8	0.020	11	0.019
14	0.915	1	0.029	1	0.019
4	0.913	4	0.029	13	0.023
10	0.902	12	0.030	10	0.023
3	0.888	3	0.031	9	0.023
7	0.870	14	0.032	12	0.023
5	0.870	7	0.035	4	0.027
12	0.868	13	0.035	7	0.028
13	0.864	2	0.036	8	0.031
9	0.851	10	0.041	2	0.034
8	0.822	5	0.065	3	0.042
6	0.796	9	0.080	6	0.075

Table 1. The ranking of the fourteen components according to the CV rate, according to the standard deviation of the CV rate across pose, and according to the standard deviation across identity.

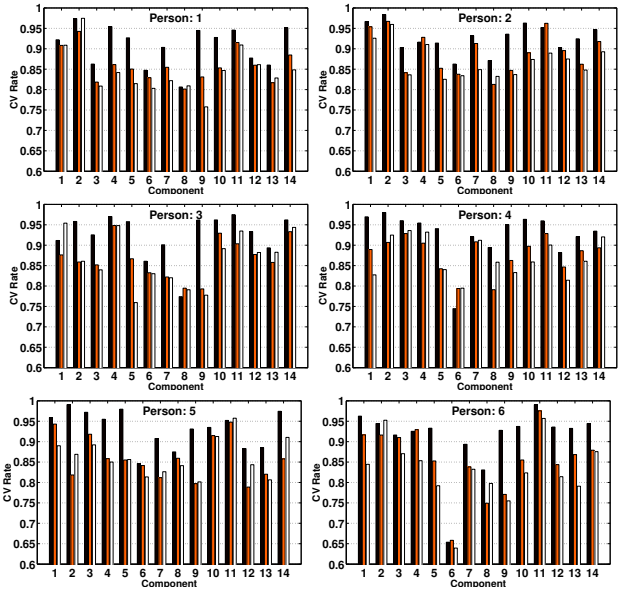

Figure 7. The CV rates of the components grouped by subjects. Each diagram shows the results for one subject. Each group of three bars represents one component across the three pose intervals, starting with pose interval 1 from the left.

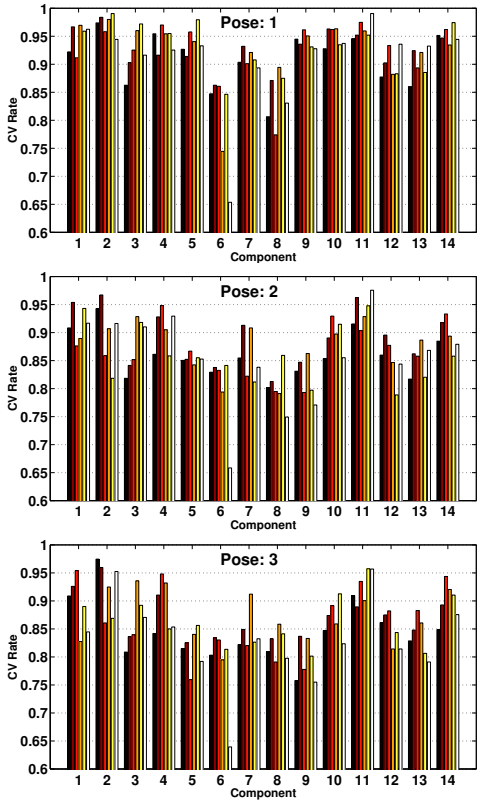

Figure 8. The CV rates of the components grouped by pose. Each diagram shows the results for a particular pose interval. Each group of six bars represents one component across the six subjects, starting with person 1 from the left.

tion is most relevant for applications where the database is relatively small, for example in a home or office environment. Especially for verification, looking at specific facial features that distinguish a given person from other people in the database seems a sensible approach. The results about the relevance of various facial parts and their dependency on the pose might generalize to larger databases and therefore be useful to a broad spectrum of component-based recognition systems. A couple of issues relevant for building a recognition application were not addressed in this paper, e.g. the problem of locating a specific component in a face. We also did not investigate the robustness of components against changes in facial expression and changes in the appearance of person's face over time. These issues might be analyzed using more sophisticated 3D morphable face models, which have the capability of modelling facial expression and aging.

5. Conclusion

We presented a method for automatically learning a set of discriminatory facial components for face recognition. The algorithm performed an iterative growing of components starting with small initial components located around preselected points in the face. The direction of growing was determined by the gradient of the cross-validation error of the component classifiers. Experiments were conducted on images of six subjects which were split into three pose intervals ranging from $0°$ to $44°$ of rotation in depth. The results show that the shape of the learned components and their cross-validation error heavily depend on both the subject and the pose, suggesting the use of pose- and subject-specific components for face recognition. Most components showed an increase of the cross-validation error with increasing rotation in depth. Averaged over all subjects and poses, the components around the eyes, the eyebrows and the chin performed the best while the component around the tip of the nose had the highest cross-validation error.

Acknowledgements

The authors would like to thank Volker Blanz for generating the 3D face models.

References

[1] D. J. Beymer. Face recognition under varying pose. A.I. Memo 1461, Center for Biological and Computational Learning, M.I.T., Cambridge, MA, 1993.

[2] V. Blanz and T. Vetter. A morphable model for synthesis of 3D faces. In *Computer Graphics Proceedings SIGGRAPH*, pages 187–194, Los Angeles, 1999.

[3] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[4] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Proc. 8th International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, 2001.

[5] B. Heisele, T. Serre, M. Pontil, T. Vetter, and T. Poggio. Categorization by learning and combining object parts. In *Neural Information Processing Systems (NIPS)*, pages 1239–1245, Vancouver, 2001.

[6] C. Wallraven, A. Schwaninger, S. Schumacher, and H. H. Bülthoff. View-based recognition of faces in man and machine: Revisiting inter-extra-ortho. *Lecture Notes in Computer Science*, LNCS 2525:651–660, 2002.

[7] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775 –779, 1997.