FACE DETECTION IN COLOR IMAGES USING PRINCIPAL COMPONENTS ANALYSIS

B. Menser and F. Müller

Aachen University of Technology (RWTH), Germany

ABSTRACT

In this paper we present a face detection algorithm for color images with complex background. We include color information into a face detection approach based on principal components analysis (PCA). A skin color probability image is generated by doing a color analysis and the PCA is performed on this new image instead of the luminance image. Experiments show that color information improves the robustness of the detection significantly.

INTRODUCTION

In the recent past there is a growing interest in image content analysis, given a large number of applications like image retrieval in databases, face recognition or content-based image coding. The automatic detection of human faces in images with complex background is an important preliminary task for these applications (see for example Chellapa et al [1]).

A problem closely related to face detection is face recognition. One of the basic approaches in face recognition is the eigenspace decomposition (e.g. Turk and Pentland [8]). The image under consideration is projected into a low dimensional feature space that is spanned by the eigenvectors of a set of test faces. For the recognition task, the resulting coefficients (principal components) are compared to those of images in the database. Principal components analysis (PCA) can also be used for the localization of a face region. An image pattern is classified as a face if its distance to the face space is smaller than a certain threshold. However, our experiments show that the background leads to a significant number of false classifications if the face region is relatively small.

For the detection of facial regions in color images, several techniques have been proposed so far, using texture, shape and color information, e.g. Sobottka and Pitas [6], Saber and Tekalp [5], Wang and Chang [9]. Due to the fact that color is the most discriminating feature of a facial region, the first step of many face detection algorithms is a pixel-based color segmentation to detect skin-colored regions. The performance of such a hierarchical system is highly dependent on the results of this initial segmentation. The subsequent classification based on shape may fail if only parts of the face are detected or the face region is merged with skin-colored background.

In this paper we incorporate color information into a face detection scheme based on principal components analysis. Instead of performing a pixel-based color segmentation, we create a new image which indicates the probability of each image pixel belonging to a skin region (skin probability image). Using the fact that the original luminance image and the probability image have similar grey-level distributions in facial regions, we employ a principal components analysis to detect facial regions in the probability image. The utilization of color information in a PCA framework results in a robust face detection even in the presence of complex and skin colored background. Our detection algorithm finds frontal views of human faces in color images over a range of scales.

COLOR ANALYSIS

The color of human skin is distinctive from the color of many other natural objects, hence color is a very important feature that can be used for face detection. Analyzing the skin-tone color statistics, one observes that skin colors are distributed over a small area in the chrominance plane and the major difference between skin tones is intensity. Thus, the image is first converted into a color space which provides a separation into a luminance channel and two chrominance components like the YCbCr color space.

Let $\boldsymbol{w}_{ij}$ denote the vector which is formed by the chrominance components of the pixel at site (i, j). The conditional probability function of $\boldsymbol{w}_{ij}$ belonging to the skin class S is modeled by a two-dimensional Gaussian

$$p(\boldsymbol{w}_{ij}|S) = \frac{\exp[-\frac{1}{2}(\boldsymbol{w}_{ij} - \boldsymbol{\mu}_S)^{\mathrm{T}}\boldsymbol{\Sigma}_S^{-1}(\boldsymbol{w}_{ij} - \boldsymbol{\mu}_S)]}{2\pi|\boldsymbol{\Sigma}_S|^{\frac{1}{2}}}. \tag{1}$$

The mean $\boldsymbol{\mu}_S$ and the covariance matrix $\boldsymbol{\Sigma}_S$ of the distribution are estimated from a training set. The gaussian approach can be verified by a nonparametric density estimation (parzen estimate).

Usually, the conditional pdf is used for a classification of the image pixels. In [5] a threshold is se-

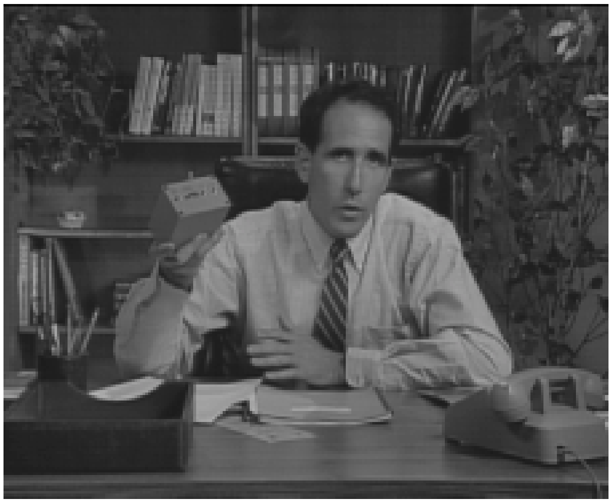


Figure 1: Luminace image (salesman#1)

Figure 2: Skin tone probability image

lected to decide whether a pixel belongs to the skin class or not. In [6] and [9] a pixel is classified as a skin-tone pixel if its chromaticity falls within a predefined region of the color space. The subsequent analysis of the face candidates is region-based and may fail if a face is not correctly extracted into one single region.
Instead of doing a binary segmentation, we preserve the continuous information given by $p(\boldsymbol{w}_{ij}|S)$. We create a so called skin probability image ϕ which indicates the probability of each image pixel belonging to the skin class S:

$$\phi(i,j) \sim p(\boldsymbol{w}_{ij}|S). \tag{2}$$

Fig. 1 and 2 shows the luminance component of the image salesman#1 and the skin probability image ϕ, respectively.
The test image has complex background with objects colored similar to human skin. The example shows that it is difficult to find a threshold such that the whole face is represented by a single region which is not connected with background objects.
An important observation is that the probability image resembles the original luminance image in facial regions. For example, parts of the face which do not have skin color like the eyes appear as dark holes in both luminance image and skin probability image. On the other hand, image regions with different color (background) are suppressed in the skin probability image. Generally speaking, the grey-level distribution of a face region is very similar in the original luminance image and the skin probability image. Since the important global features of a face (e.g. shape, eyes, mouth) are preserved, we suggest to apply a face detection algorithm based on principal components analysis to the skin probability image instead of the luminance image. The major benefit of using the skin probability image is that a facial region is enhanced compared to the background. Furthermore, the influence of different lighting conditions is drastically reduced since the skin probability depends not on the luminance.
A similar approach has been presented by Dai and Nakano [2], where a face texture model originally developed for grey-level images is applied to the I-component of an image in the YIQ color space.

PRINCIPAL COMPONENT ANALYSIS

Most face detection schemes can be divided into two different strategies. The first method is based on the detection of facial features (e.g. Yow and Cippola [10]), whereas the second approach tries to detect a face pattern as a whole unit (e.g. Sung and Poggio [7], Moghaddam and Pentland [4]). Following the second approach, each image pattern of dimension I by J can be considered as a vector $\boldsymbol{x}$ in a $N = IJ$ dimensional space. Obviously, images of faces will not be randomly distributed in this high dimensional image space. A suitable mean to reduce the dimensionality of the data set is the principal components analysis (PCA).
The central idea of principal components analysis is to find a low dimensional subspace (the feature space) which captures most of the variation within the data set and therefore allows the best least-square approximation (Jolliffe [3]). When used for face detection and recognition, this principal subspace is often called the "face space" which is spanned by the "eigenfaces" [8].
Given a set of training vectors $\{\boldsymbol{x}\}$ (face samples) with sample covariance matrix $\boldsymbol{\Sigma}$, the KLT basis can be computed by solving the eigenvalue problem

$$\boldsymbol{\Lambda} = \boldsymbol{P}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{P}, \tag{3}$$

where $\boldsymbol{P}$ is the eigenvector matrix of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ the diagonal matrix of the eigenvalues.

The orthogonal projection matrix $\boldsymbol{P}_M$ into the M-dimensional principal subspace ($M \ll N$) is given by the M eigenvectors corresponding to the largest eigenvalues. These eigenvectors ("eigenfaces") form the columns of the projection matrix $\boldsymbol{P}_M$.
The principal components vector $\boldsymbol{y}$ is obtained by projecting the image $\boldsymbol{x}$ into the face space:

$$\boldsymbol{y} = \boldsymbol{P}_M^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}), \tag{4}$$

where $\bar{\boldsymbol{x}}$ denotes the mean face image.
The classification of the image pattern containing N pixels as "face" or "non-face" is only based on its M principal components y_i.
The approximation of a face using M "eigenfaces" is given by $\hat{\boldsymbol{x}} = \boldsymbol{P}_M\, \boldsymbol{y} + \bar{\boldsymbol{x}}$. The residual reconstruction error ϵ^2

$$\begin{aligned} \epsilon^2 &= ||(\boldsymbol{x} - \hat{\boldsymbol{x}})||^2 \\ &= ||(\boldsymbol{x} - \bar{\boldsymbol{x}}) - \boldsymbol{P}_M\, \boldsymbol{y}||^2 \\ &= ||\boldsymbol{x} - \bar{\boldsymbol{x}}||^2 - \sum_{i=1}^{M} y_i^2 \end{aligned} \tag{5}$$

indicates how well the test pattern can be approximated in the face space. Thus, the "distance from face space" (DFFS) defined by equation (5) can be used to determine if the image pattern represents a face [8].
The distance between the projected image and the mean face image in the feature space is given by the norm of the principal component vector. Since the variance of a principal component y_i is given by the associated eigenvalue λ_i, the squared Mahalanobis distance d^2 provides a suitable measure of the difference between the projection of the test image and the mean face:

$$d^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{y} = \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i} \tag{6}$$

For the detection task, it is customary to use only the DFFS for the decision criterion [8]. As shown in the next section, the incorporation of the "distance in face space" (DIFS) defined by equation (6) improves the robustness of the detection significantly if the detection is performed on the skin probability image.

FACE DETECTION USING PCA

In our experiments we use a 10-dimensional principal subspace ($M = 10$). The projection matrix $\boldsymbol{P}_M$ is calculated by an eigenspace decomposition using a set of luminance face test images. To localize a facial region in a new image, the error criterion must be computed for each pixel position, resulting in a distance map. For the moment, we use a simple detector which is based only on the "distance from face space" (DFFS). The global minimum of the distance map is then selected as the best match. When using M eigenvectors, the computation of the DFFS requires $M+1$ correlations (with the M "eigenfaces" and the mean image) and an additional energy calculation. The correlations can be efficiently computed using the FFT.
First we analyze the PCA-based face detection on grey level images and point out the problems caused by the image background.

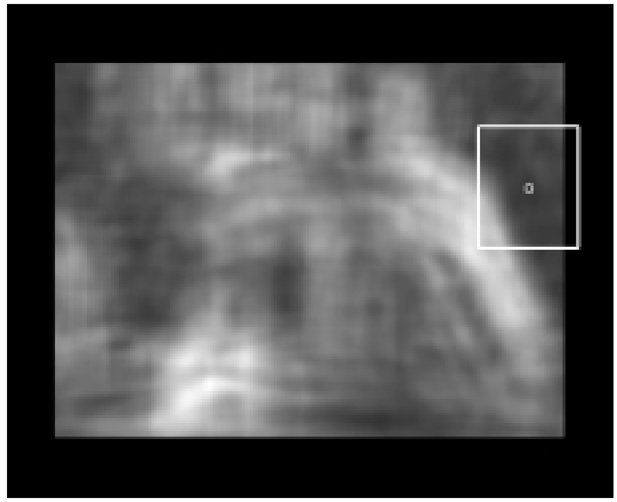
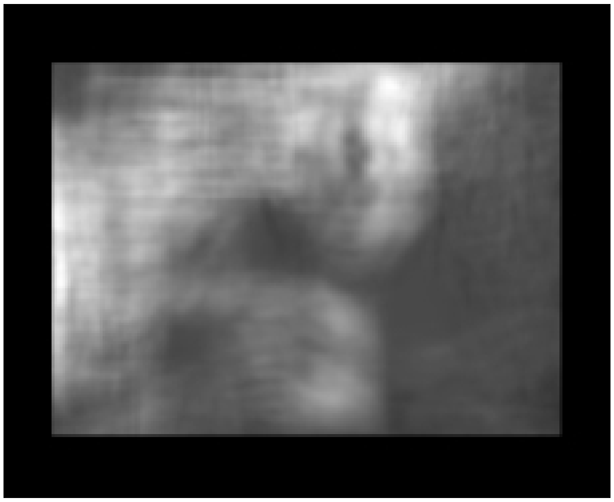
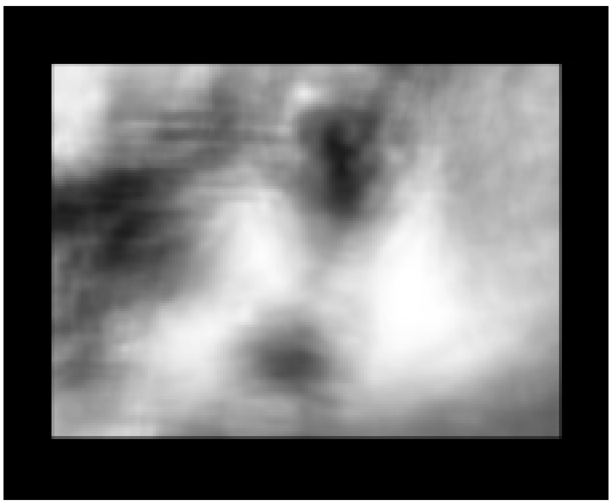


Figure 3: Distance map for the luminance component using DFFS (salesman#1)

Fig. 3 shows the DFFS map for the luminance image salesman#1. The global minimum is marked by the white circle. Though there is a local minimum at the true face position, the best match is in the background region leading to a false detection. The DFFS is high for non-facial image regions with a high changing in the intensity (e.g. the area that includes parts of the light-colored shirt and the dark background). On the other hand, the DFFS becomes relative small in non-facial regions with little variance in the intensity like parts of the background at the right side of the test image. This is due to the fact that an image pattern that can be modeled by noise can be better represented by the eigenfaces than a non-facial image pattern containing a strong edge. Therefore, detection based on the DFFS becomes difficult even in images with a simple background if the face region does not cover the main part of the test image.
We now apply a principal component analysis to the skin probability image defined by equation (2) using the same projection matrix. The DFFS map for the skin probability image shown in fig. 2 is displayed in fig. 4a. The true face region is characterized by a local minimum. Similar to the luminance case, the error criterion is also low in background regions with little variance. These regions represent non-skin-colored background, so the probability image is near zero in these areas.
Fig. 4b shows the "distance *in* face space" for each

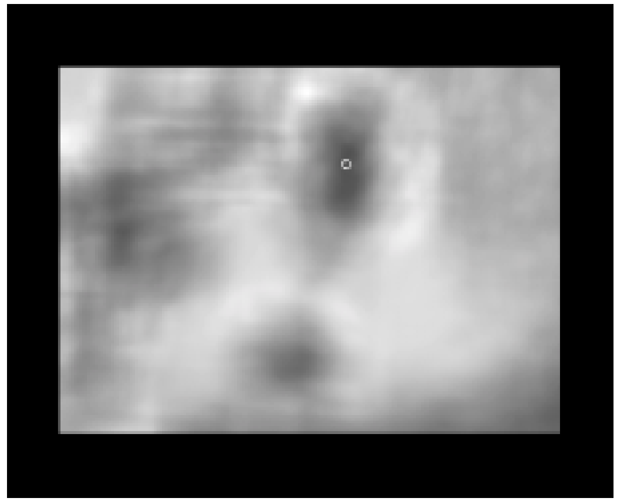


a) DFFS



b) DIFS

Figure 4: Distance maps for the skin probability image (salesman#1)



a) Combined error criterion

b) Detected face region

Figure 5: Face detection using DFFS and DIFS

spatial position which is defined by the weighted LS-norm of the principal component vector. The projection of an image region with values near zero into the face space results in large principal components. Thus, the DIFS can be employed to eliminate the influence of the non-skin-colored background. Since no additional correlation is necessary for the calculation of the DIFS, the computational load increases only slightly. We define a new error criterion which is a weighted sum of the DFFS and the DIFS:

$$e = d^2 + c\,\epsilon^2 \tag{7}$$

Since the principal components are scaled by the corresponding eigenvalues for the calculation of the DIFS, a similar scaling is necessary for the DFFS. Thus, we set the scaling factor to $c = \frac{1}{k\,\lambda_M}$ where λ_M denotes the smallest calculated eigenvalue and k is a small constant ($k = 1...5$) determined by experiments.

In [4] the eigenspace decomposition is used to estimate the complete probability density for a certain object class (e.g. the face class). Assuming a normal distribution, the estimate for the Mahalanobis distance is also a weighted sum of the DFFS and the DIFS similar to equation (7). A main difference is that our scaling factor c is much smaller than the one used in [4] because the DIFS provides more information when using the skin probability image instead of the luminance image. The distance map using the combined error criterion is shown in fig. 5a. The global minimum (marked by the white circle) lies in the true face region and the face is detected correctly. Fig. 5b shows the detected face region superimposed on the luminance component.

To detect faces at multiple scales, the detection algorithm is performed on several scaled versions of the skin probability image and the global minimum of the resulting multi-scale distance maps is selected

Figure 6: Multiple scale face detection

as the face position. Detection results for several images of MPEG test sequences are shown in fig. 6.

The global minimum detection based on the assumption that the image contains exactly one face. To detect several faces and reject non-facial images, a threshold can be introduced which allows a trade-off between false detection and false rejection. If the error criterion at a certain spatial position is less than this predefined threshold, the subimage located at this position is classified as a face. To prevent overlapping regions, only the global minimum is selected in the first step. For the search of the second (local) minimum, all spatial positions which lead to overlapping regions are discarded. This procedure is repeated until no position with error less than the threshold can be found which is not already covered by another detected region. The result of a multiple face detection is given in fig. 7.

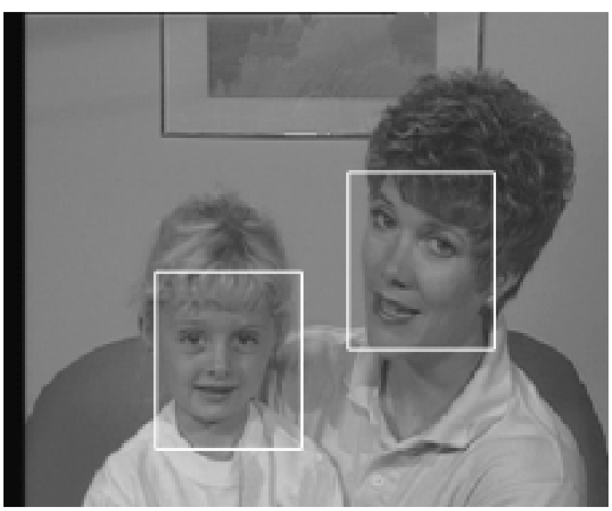


Figure 7: Detection of more than one face

CONCLUSION

The presented face detection algorithm combines a skin color analysis with an eigenspace approach. The incorporation of color information reduces the influence of the image background and improves the detection performance especially in images where the size of the facial area is relatively small. Our current work is focused on the adjustment of the error criterion and the automatic adaption of the global detection threshold. Furthermore, the algorithm will be extended to face tracking in image sequences.

REFERENCES

[1] Chellapa R., Wilson C. L. and Sirohey S., 1995, “Human and machine recognition of faces: A survey”, Proc. IEEE, 83(5), 705–740

[2] Dai Y. and Nakano Y., 1996, “Face-texture model based on SGLD and its application in face detection in a color scene”, Pattern Recognition, 29(6), 1007–1017

[3] Jolliffe I. T., 1986, “Principal Component Analysis”, Springer, New York, Berlin, Heidelberg, Tokyo

[4] Moghaddam B. and Pentland A., 1997 “Probabilistic visual learning for object representation”, IEEE Trans. PAMI, 19(7), 696–710

[5] Saber A. and Tekalp A. M., 1998, “Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions”, Pattern Recognition Letters, 19(8), 669–680

[6] Sobottka K. and Pitas I., 1998 “A novel method for automatic face segmentation, facial feature extraction and tracking”, Signal Processing: Image Com., 12(3), 263–281

[7] Sung K. K. and Poggio T., 1998, “Example-based learning for view-based human face detection”, IEEE Trans. PAMI, 20(1), 39–51

[8] Turk M. and Pentland A., 1991, “Eigenspaces for recognition”, Journal of Cognitive Neuroscience, 3(1), 71–86

[9] Wang H. and Chang S. F., 1997, “A highly efficient system for automatic face region detection in MPEG video”, IEEE Trans. Circ. Syst. Video Tech., 7(4), 615–628

[10] Yow K. C. and Cipolla R., 1997, “Feature-based human face detection”, Image and Vision Computing, 15(2), 713–735