face detection and recognition,
facial feature extraction, face masks

Michal KAWULOK[*], Bogdan SMOLKA[**]

APPLICATION OF COLOR INFORMATION IN HUMAN FACE RECOGNITION

This paper presents methods of utilizing color data in automatic human face recognition. An existing approach to color-based face detection is described. A new concept of using color information for feature extraction improvement is proposed. By utilizing color data, face maps are generated and transformed into masks of human face regions weights. The new approach has been tested and the results are presented.

1. INTRODUCTION

Human face recognition [2, 10, 12] is one of the most popular biometric techniques. It is characterized by a low level of required interaction with a person being recognized, but offers relatively low effectiveness comparing to other biometric methods [1, 7]. There are many applications of face recognition, ranging from entertainment to access control and surveillance tracking.

It is worth noticing that the majority of face recognition algorithms process grayscale images. It can be considered as an advantage, especially when only such images are available. However, this also means that application of color, high resolution images, which nowadays are often available, would not improve the face recognition effectiveness using the existing techniques.

2. COLOR-BASED FACE DETECTION

An approach proposed by Hsu et al. [3] takes advantage of selected common properties of human skin color. It was observed that after transforming an image from the *RGB* space to the YC_bC_r space, it is relatively easy to detect a face and facial features with high effectiveness.

The first step is the lighting compensation and skin-tone pixels detection based on an elliptical skin model in the nonlinearly transformed YC_bC_r space. It is possible to define a skin color in this way and detect skin areas. When a face is detected, eyes and mouths are found by constructing maps in the YC_bC_r color space (Fig. 1). Eyes are characterized by high blue and low red intensities, as well as by many dark and bright pixels. These facts may be utilized for designing the eye map *EM*:

[*] Silesian University of Technology, Institute of Computer Science, 16 Akademicka St., 44-101 Gliwice, Poland
[**] Silesian University of Technology, Institute of Automatic Control, 16 Akademicka St., 44-101 Gliwice, Poland

$$EM = (EM_C)\cdot(EM_L)\,,\ EM_C = \frac{1}{3}(C_b^{\,2} + \overline{C}_r^{\,2} + \frac{C_b}{C_r})\,,\ \overline{C}_r = 255 - C_r\,,\ EM_L = \frac{Y_{dil}(p,q)}{Y_{er}(p,q)+1}\,, \tag{1}$$

where EM_C is a chrominance map, EM_L is a luminance map and EM is the final eye map. Y_{dil} and Y_{er} is the luminance channel, dilated and eroded respectively, p and q are pixel coordinates. Mouth pixels contain higher C_r values and lower C_b values comparing to other face regions. This observation makes it possible to construct a mouth map:

$$MM = C_r^{\,2}\cdot\left(C_r^{\,2} - \eta\cdot\frac{C_r}{C_b}\right)^2,\ \eta = 0.95\cdot\frac{\mathrm{avg}(C_r^{\,2})}{\mathrm{avg}(C_r/C_b)}\,, \tag{2}$$

where η is a ratio of the average C_r^2 value (avg(C_r)) to the average C_r/C_b value (avg(C_r/C_b)). Finally, the maps are eroded, dilated and normalized in order to eliminate noise information, which is usually weaker than the response of eye and mouth regions.

Fig. 1. From the left: face image (M_023_1 from the AR database), Y, C_b and C_r channels, eye map and mouth map.

By analyzing the described maps, an exact location of feature points may be obtained, which means that using color information only, it is possible to detect faces and facial feature points.

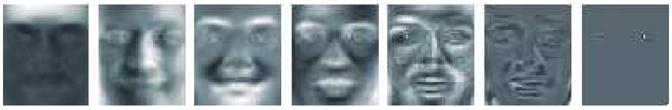
3. FEATURE EXTRACTION

In the case of face detection, common properties of face and skin color have been utilized [3]. However, in the case of feature extraction it is desirable to find such properties which are characterized by high variance among images belonging to different people and low variance for images of the same person.

3.1 THE EIGENFACES METHOD

One of the most popular face feature extraction methods is the Eigenfaces method [9] which utilizes the Principal Component Analysis (PCA). This method consists of two stages. The first one is the training stage. It is performed for N sample face images, which have been normalized in such a way that they are of equal size and eyes are located in fixed positions. Using these images, a covariance matrix ($\mathbf{C} = 1/N\sum_{i=1}^{N}(\mathbf{x_i}-\boldsymbol{\mu})(\mathbf{x_i}-\boldsymbol{\mu})^{\mathrm{T}}$) is built, where $\mathbf{x_i}$ is the i-th normalized face image vector of a constant length l and $\boldsymbol{\mu}$ is an average normalized face image vector. Subsequently, the eigenvectors ($\mathbf{u}$) of the covariance matrix are calculated and they are sorted in descending order by corresponding eigenvalues (λ). Because a face space has far less dimensions than the input space, a relatively small number (l') of the eigenvectors with the highest eigenvalues creates an orthogonal basis for the l'-dimensional face space. If the eigenvectors are scaled to the pixels value range, they are similar to faces in appearance and they are called eigenfaces (see Fig. 2). For the tested training set relevant information is concentrated in the first 150 eigenfaces and the remaining ones contain mainly noise, so the face space in this case can be described by approximately 150 dimensions.

When the training is completed, features can be extracted from any normalized face image, utilizing the calculated l' eigenvectors with the highest eigenvalues. The feature extraction is



1 3 5 10 30 80 160

Fig. 2. Examples of eigenfaces sorted by the eigenvalues in descending order (given under each image).

performed by calculating a dot product between a normalized image, which features are being extracted, and every eigenvector which defines the face space ($w_i = \mathbf{u_i} \cdot \mathbf{x}$, i=1..l', where w_i is the i-th element of a feature vector, $\mathbf{u_i}$ is the i-th eigenvector and $\mathbf{x}$ is a normalized image). In this way a face is described by a feature vector of length equal to l'.

3.2 IMPROVEMENTS TO THE EIGENFACES METHOD

It can be observed that the Eigenfaces method does not take advantage of face topology and each pixel is taken with equal weight. It is a serious drawback, because characteristic data that make it possible to distinguish between two faces are not equally distributed over a face. Intuitively, areas near eyes and nose carry more discriminative information than other parts of a face and have extra-personal nature. They are characterized by high variance within the training set, which affects the covariance matrix. As a result, the eigenvectors that show the main directions of variance among the images enhance these discriminative differences. Hence, in the case of two face images with different eyes and noses, derived from two different persons, the feature vectors will differ significantly too. The differences concern also other face areas, for example mouth and cheeks, which leads to undesired consequences, because two images of the same person often differ in these regions (e.g. due to face expression changes). Therefore, these regions have intra-personal nature.

This disadvantage can be overcome by applying face relevance masks that emphasize importance of extra-personal face regions and decrease influence of the intra-personal ones [4]. A feature vector can be calculated including a mask vector $\mathbf{M}$: $w_i = \sum_{j=1}^{l} u_{ij} \cdot x_j \cdot M_j$, where l is a number of image pixels and u_{ij}, x_j, M_j are the j-th elements of the vectors: i-th eigenvector, normalized face image and mask. Face masks are described in [4] and some examples of them are presented in Fig. 3. The best results are given by a difference mask, based on statistical differences between images: areas which differ more between images belonging to different classes induce higher values and those, which differ between images belonging to the same class, induce lower values.It must be noted, that these masks were obtained taking into account grayscale image information only.

3.3 MASK FROM COLOR

It is quite difficult to find discriminative information specific to color that is not present in grayscale images. However, the experiments show that it is possible to extract information from color data which help to define importance of image areas to create better masks than those based only on grayscale data. The first possibility is to generate a mask directly from face maps used for detection purposes. In this case the eye map should increase the mask values, whereas the mouth map should be subtracted from the mask. This operation rises an influence of areas near eyes and decrease an importance of mouth region which is considered as having intra-personal nature (see Fig. 4). The maps were dilated and blurred in order to cover larger face areas.

Fig. 3. Examples of face masks. From the left: a manually created "T" mask and two difference masks.

A different approach is to calculate the exact facial feature positions and generate a mask taking into account the set of detected points. The maps extracted from color data make it possible to obtain exact positions of eyes and mouth and to estimate the nose position as well. Each point can be treated as a source of Gaussian distribution of weights, which create a submask. Hence, the final mask is a sum of submasks obtained from all the feature points. The submask intensity ω generated by the point (p_0, q_0) can be calculated for every point (p, q):

$$\omega_{p_0,q_0}(p,q) = \omega_{\max} \cdot \exp\left[-\frac{d_{p_0,q_0}{}^2(p,q)}{\sigma^2}\right],\ d_{p_0,q_0}(p,q) = \sqrt{\delta_h \cdot (p-p_0)^2 + \delta_v \cdot (q-q_0)^2} \tag{3}$$

where ω_{max} is the submask maximal value, σ is a width parameter and $d_{p0,q0}$ is a weighted distance from the source point. The vertical and horizontal distances are assigned with separate weights (δ_v for vertical and δ_h for horizontal direction) in order to model elliptical submasks, which are better suited to the properties of human face. For example, when $\delta_v > \delta_h$, a flat ellipse is created, which is very useful for the mouth region submask generation. The submask may also have several sources (for example one submask for both eyes and nose) and its intensity can be calculated in this way:

$$\omega(p,q) = \omega_{\max} \cdot \exp\left[-\frac{1}{\sigma^2}\left(\sum_i d_{xi,yi}(p,q)\right)^2\right]. \tag{4}$$

Depending on the parameters, as well as on a number and types of the submasks, different final masks can be obtained using this technique, examples of which are shown in Fig. 5. The described masks obtained from colors can be imposed on normalized images in order to change individual pixel values and their influence on the final feature vector values (see Fig. 4). An important advantage of this approach is that a mask is generated for every image separately and it is possible to fit the mask to a face better and enhance the extra-personal areas with higher precision.

4. EXPERIMENTS

The experiments were performed on two sets of images. The first one consists of 1136 color images from the AR database [5] and the second one contains 5669 color images from the Notre-Dame database [6]. Each set was divided into a training set which contains one image per person and a query set which contains images of people whose images were included in the training set. The training set contains 136 images in the case of the AR database and 275 for the Notre-Dame database. Number of images in the query set was 1000 and 5394, respectively. Face recognition effectiveness was measured for the images from the query set. Every image from the query set was compared with all the images from the training set. This single recognition was successful, if the similarity rate between the tested image and an image from the training set, which contained a face of the same person as the tested image, was the highest for all the images in the training set.

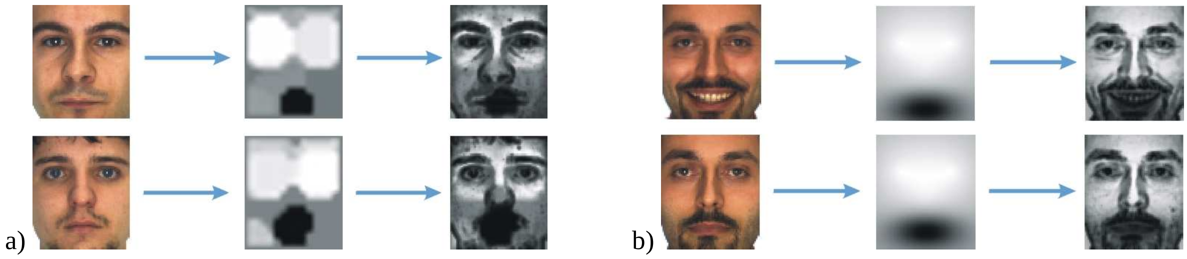


Fig. 4. Examples of face masks obtained directly from eye and mouth maps (a: M_001_1 and M_003_1 images from AR database [5]) and generated basing on a set of detected feature points (b: M_004_1 and M_004_15 images).

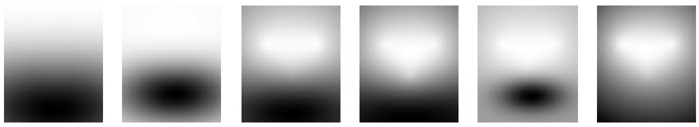


Fig. 5. Examples of face masks based on a set of detected feature points generated for various parameters.

The recognition effectiveness was tested for the Eigenfaces method with no improvements, with the application of difference masks and masks from color. The results are presented in Table 1 and the cumulative match characteristics [7, 8] are shown in Fig. 6. During the experiments various versions of difference masks and masks from color were applied. In the case of the masks from color, various sets of parameters that define the feature point submasks were used. The best cases of these two types of masks are presented and compared in this paper. The best results were achieved for a mask based on a set of detected points, which was created as a sum of four intensity submasks: left and right eye submask ($\omega_{max} = 3$, $\delta_h = \delta_v = 1$, $\sigma = 7$), mouth submask ($\omega_{max} = 1$, $\delta_h = 1$, $\delta_v = 4$, $\sigma = 44.7$) and a submask with three sources (both eyes and nose, $\omega_{max} = 3$, $\delta_h = \delta_v = 1$, $\sigma = 4.5$). The mouth submask was a negative one, whereas the remaining ones were positive.

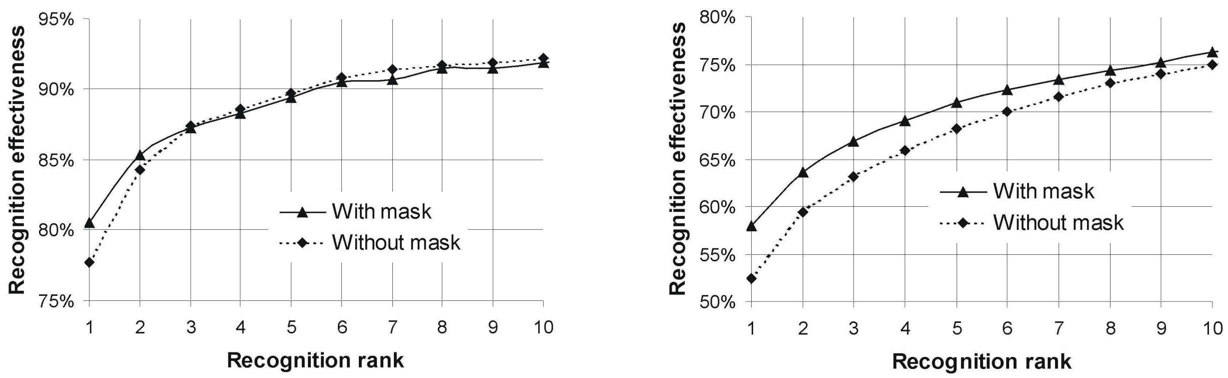


Fig. 6. Cumulative match score for the AR database (left) and for the Notre-Dame database (right).

The experiments have shown that for both large sets of face images the application of face masks improved the results and more faces were correctly recognized. The effect of the error rate reduction is more visible in the case of the Notre-Dame database, which contains more difficult cases and is characterized by lower recognition rates. It is worth noticing, that the relative reduction of the error rate after applying the color-based masks is similar in the case of both databases.

Method	Error rate	
	AR	Notre-Dame
Eigenfaces	22.3 %	47.5 %
Eigenfaces with difference mask	20.6 %	43.2 %
Eigenfaces with mask from color	19.5 %	42.0 %
Error reduction (relative change)	12.6 %	11.6 %

Table 1. Recognition error rates.

5. CONCLUSIONS AND FUTURE WORK

The presented approach utilizes an idea of improving face recognition effectiveness by taking into account color data. The experiments have proved that it gives better results than the pure Eigenfaces method. In the case of several thousands of color images from two different databases the error rate has been reduced by over 10%. The main conclusion is that color data may be treated as a reliable source of information which describes the nature of human face features. Considering that information it is possible to increase influence of discriminative extra-personal features and to reduce the undesired influence of intra-personal ones.

The future work will be concentrated on finding better ways of extracting additional information from color data and using it for feature extraction. It is also planned to investigate the possibilities of applying this approach to other methods of human face feature extraction (e.g. the Elastic Bunch Graph Matching method [11]).

BIBLIOGRAPHY

[1] GONG S., MCKENNA S.J., PSARROU A., *Dynamic Vision From Images to Face Recognition*, Imperial College Press, 1999

[2] GROTHER P., *Face Recognition Vendor Test 2002, Supplemental Report*, February 2004, NIST IR 7083

[3] HSU R.-L., ABDEL-MOTTALEB M., JAIN A.K., “Face detection in color images”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, No. 5, May 2002, pp. 696-706

[4] KAWULOK M., “Masks and Eigenvectors Weights for Eigenfaces Method Improvement”, *Proceedings of Computer Vision and Graphics International Conference, ICCVG 2004*, Springer 2006, pp. 528-533

[5] MARTINEZ A.M., BENAVENTE R., “The AR Face Database”, *CVC Technical Report #24*, June 1998, http://rvl1.ecn.purdue.edu/~aleix/ar.html

[6] PHILLIPS P.J., FLYNN P.J., SCRUGGS T., BOWYER K.W., CHANG J., HOFFMAN K., MARQUES J., J. MIN J., WOREK W., “Overview of the Face Recognition Grand Challenge”, *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005

[7] PHILLIPS P.J., GROTHER P.J., MICHEALS R.J., BLACKBURN D.M., TABASSI E., BONE J.M., *Face recognition vendor test 2002: Evaluation report*, NISTIR 6965, 2003, available online at http://www.frvt.org

[8] PHILLIPS P.J., WECHSLER H., HUANG J., RAUSS P., “The FERET database and evaluation procedure for face recognition algorithms”, *Image and Vision Computing J*, Vol. 16, No. 5, 1998, pp 295-306

[9] TURK M., PENTLAND A., “Face Recognition Using Eigenfaces”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1991*, pp. 586-591

[10] WECHSLER H., PHILLIPS P.J., BRUCE V., SOULIE F.F., HUANG T.S., *Face Recognition: From Theory to Applications*, Eds. Springer-Verlag, Berlin, Germany, 1998

[11] WISKOTT L., FELLOUS J.M., KRUGER N., MALSBURG C., “Face Recognition by Elastic Bunch Graph Matching”, *Tech. Report 96-08*, Ruhr-Universitat Bochum, April 1996

[12] ZHAO W., CHELLAPPA R., PHILLIPS P.J., ROSENFELD A., *Face Recognition: A Literature Survey, Technical Report CARTR-948*, Center for Automation Research, University of Maryland, College Park, MD, 2000