ELSEVIER

Pattern Recognition Letters 22 (2001) 1525–1536

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Mixture model for face-color modeling and segmentation

Hayit Greenspan [a,*], Jacob Goldberger [b,1], Itay Eshet [a]

[a] Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel-Aviv 69978, Israel
[b] The Weizmann Institute of Science, Rehovot 76100, Israel

Received 18 July 2000; received in revised form 28 March 2001

Abstract

In this paper, we propose a general methodology for face-color modeling and segmentation. One of the major difficulties in face detection and retrieval is partial face extraction due to highlights, shadows and lighting variations. We show that a mixture-of-Gaussians modeling of the color space, provides a robust representation that can accommodate large color variations, as well as highlights and shadows. Our method enables to segment within-face regions, and associate semantic meaning to them, and provides statistical analysis and evaluation of the dominant variability within a given archive.

Keywords: Face-color modeling; Skin color modeling; Face segmentation; Gaussian mixture

1. Introduction

Automatic detection and localization of human faces is a difficult task. In recent years, an increasing body of research has addressed this field, in developing algorithms to detect faces in 'mug-shot' images, or detect faces in uncontrolled, complex background, environments (Phillips et al., 1998). Face detection has important applications, as a first step in higher-level face recognition tasks, such as personal identification for security purposes, and in the field of multimedia, for face retrieval in video indexing applications or for interactive human–machine interfaces.

Existing methods may be roughly divided into several categories: local facial feature detection, feature-based (color, texture, shape and motion), and template-based using pattern recognition techniques. In the first category, algorithms are used to detect facial features such as eyes and mouth. Statistical models of faces are used to verify the face hypothesis (e.g., Yang and Huang, 1994; Chang et al., 1994; Burl et al., 1995; Yow and Cipolla, 1997; Jeng et al., 1998). Feature-based approaches use one or a combination of features to detect face regions in an image. Included are color-based schemes (which will be elaborated on in this paper), texture-based methods, shape-based methods (such as edge-based schemes and active contours), as well as motion-based schemes that may detect the face region as a moving object in an image sequence. These features may be useful as a preprocessing step in which candidate face regions are extracted, from which facial features may be found; alternatively,

* Corresponding author. Tel.: +972-3-6407398; fax: +972-3-6407939.
E-mail address: hayit@eng.tau.ac.il (H. Greenspan).
[1] Currently with CUTe Systems Ltd., Tel Aviv, Israel.

0167-8655/01/$ - see front matter
PII: S 0 1 6 7 - 8 6 5 5 (0 1) 0 0 0 8 6 - 1

they may serve as components in a characteristic (learned) feature-vector for the face class. In the third category, correlation templates are used to detect local sub-features that can be rigid in appearance such as view-based eigenfaces (Turk and Pentland, 1991), or deformable (deformable templates, Wiskott et al., 1997). Alternative pattern-based approaches are based on neural-networks (e.g., Lin et al., 1997; Rowly et al., 1998; Sung and Poggio, 1998). The advantages are that the modeling and classification rules are learned from a given collection of representative face images, with no human-designed rules. The drawbacks include computational complexity and the difficulty in characterizing the 'non-face' category.

In the last several years, an increasing body of research has addressed the specific problem of automatic face detection based on skin color (e.g., Dai and Nakano, 1995; Saber et al., 1996; Sobottka and Pitas, 1996; Sun et al., 1998; Kim et al., 1998; Terrillon and Akamatsu, 1998; Garcia and Tziritas, 1999; Wu et al., 1999). Color is a powerful fundamental feature that can be used as a first step in the process of face detection in complex scenes. Color image segmentation is computationally efficient. It is invariant to the viewpoint or scale of a given face and is robust to cluttered backgrounds. Thus, color information may be utilized to locate candidate face regions, following which additional face features may be used in order to validate the face hypothesis. Shape constraints, such as ellipsoids and symmetry constraints may be used, as well as facial features extracted and their relative geometrical ordering used for testing the face hypothesis (e.g., Chang et al., 1994; Lee et al., 1996; Wei and Sethi, 1999; Wang and Sung, 1999).

The color representation of a face obtained by a camera is influenced by many factors such as ambient light and object movement. Different cameras may produce significantly different color values, even for the same person under the same lighting condition. Finally, there exists natural variability in face color between any two people, let alone any two ethnic groups. Several recent works have focused on comparing skin color modeling at a variety of color spaces (e.g., Terrillon and Akamatsu, 1998; Garcia and Tziritas, 1999; Wu et al., 1999). Experience shows that regardless of the color space chosen and the specific modeling scheme, there are two major problems with available face segmentation schemes. First, highlights and shadows on parts of a face usually result in partial face extraction. The second issue relates to background colors that are close in color space to skin color, and which come through any face-color filtering. Any such regions are treated as candidate face regions. Shape constraints can be incorporated to facilitate solutions to the second problem. In this work, we focus on a solution to the first. We suggest a model that can accommodate extreme variations within a face database or face archive; specifically those variations relating to extreme lighting situations.

A unimodal elliptical Gaussian probability density function (p.d.f.) is most frequently used to model skin color. The overall approach is to try and find the better color space in which a unimodal Gaussian distribution accommodates the sample distribution. In this work, we suggest a different approach. Rather than trying to find the best space, lets find the right modal. Rather than trying to use simplified, single mode p.d.f., our approach is to utilize unsupervised, multi-mode semi-parametric modeling, to cluster variations of natural color, as well as variations in the lighting conditions. It is our hypothesis that with sufficient descriptive power, multiple variations may be modeled, eliminating the need for major pre-processing and/or post-processing steps. We do not aim at color invariance, but rather we aim at unsupervised clustering of the color space into a set of representative clusters, with the clusters reflecting the variations in the space.

In a few recent works, a mixture model is suggested (e.g., McKenna et al., 1998; Yang and Ahuja, 1999). In these works, the mixture model is used as a purely statistical modeling tool, with no in-depth analysis of the actual model learned. No semantic meaning is associated with the extracted clusters. We have also not found any correspondence made to the lighting variability problem. Updating and evolving the color distribution model under varying illumination, in *image sequences*, is a topic of yet another set of works (e.g., Yang et al., 1998; Sigal et al., 2000). It is clear that if available and relevant to the data in hand, temporal information can be used to improve the accuracy of detection and segmentation.

In this work, we are assuming a stationary domain in which an archive of images is provided to the system. We are thus not including temporal information. Our main objective is to take the probabilistic modeling approach further, analyzing the variability of the input space via the extracted model, and utilizing the suggested methodology for more informative segmentation. An important benefit of the proposed approach is the ability to segment within-face regions, and associate semantic meaning to them.

The paper is organized as follows. In Section 2, we focus on face-color modeling via a Gaussian mixture model. After a representative model has been extracted, we may use it in several ways. Two ways of utilizing the extracted information content are presented in this paper. Face segmentation is discussed in Section 3. A statistical analysis of the database content is shown in Section 4. Section 5 concludes the paper with discussion and future directions.

2. Modeling face color

We propose to use a Gaussian mixture model to model the face-color distribution in a selected color space. In this paper, we choose to work with the normalized r–g color space which is frequently found in the face research literature. The presented modeling scheme is a general one and may be applied to any chosen color space (e.g., HSV, CIE-xy, YIQ, Luv, Lab and TSL).

Face-color modeling is a one time global learning process per given archive. A large sample set of face-color pixels, the training set, is required for learning the model. We generate the training sets in a pre-processing phase, in which sample face pixels are collected from given training databases. In this work, we focus on heterogeneous databases that include skin tone variations as well as natural lighting variations.

In order to motivate the proposed model we start by taking a closer look at face sample distributions in color space. We use a large set of sample points from a heterogeneous database of faces. Fig. 1 shows a set of 2D histograms representing face-color samples from the ARH (a) and ARL (b) archives (see definition of archives in Section 2.2). A unimodal distribution is present in the ARH database. In the ARL database, faces are lighted selectively, thus shadows and highlights are strongly present. It is evident that a single Gaussian cannot model the two major peaks in the distribution. The need for multi-modal p.d.f. is clearly motivated.

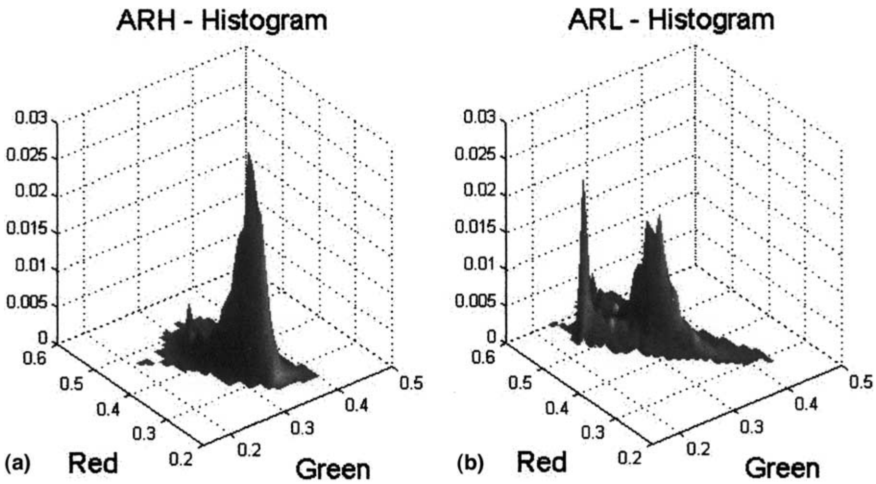

Fig. 1. 2D histograms of face-color samples (r–g space). (a) Histogram of face-color samples in a homogeneous color dataset (ARH). A strong unimodal distribution is evident. (b) Histogram of face-color samples with non-homogeneous lighting (ARL). The distribution is a much more complex multi-modal distribution.

2.1. Maximum likelihood face-color modeling via a Gaussian mixture model

We propose to model the face-color distribution using a mixture-of-Gaussians distribution. Gaussian mixtures are semi-parametric models. Maximum likelihood estimation provides means for optimizing the mixture parameters. The Expectation Maximization (EM) algorithm is used to extract the maximum likelihood estimate of the Gaussian mixture model (Dempster et al., 1997).

We start with the general estimation model. The distribution of a random variable $X \in \mathbb{R}^d$ is a mixture of k Gaussians if its density function is

$$f(x \mid \theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \times \exp\left\{ -\frac{1}{2}(x-\mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x-\mu_j) \right\} \quad (1)$$

such that the parameter set $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$ consists of:

- $\alpha_j > 0$, $\sum_{j=1}^{k} \alpha_j = 1$;
- $\mu_j \in \mathbb{R}^d$ and Σ_j is a $d \times d$ positive definite matrix.

Given a set of color vectors $x_1, \ldots, x_n$, the maximum likelihood estimation of θ is

$$\theta_{\mathrm{ML}} = \arg\max_{\theta} f(x_1, \ldots, x_n \mid \theta). \quad (2)$$

The EM algorithm is an iterative method to obtain θ_{ML}. Given the current estimation of the parameter set θ, each iteration of the EM algorithm reestimates the parameter set according to the following two steps:

- Expectation step:

$$w_{tj} = \frac{\alpha_j f(x_t \mid \mu_j, \Sigma_j)}{\sum_{i=1}^{k} \alpha_i f(x_t \mid \mu_i, \Sigma_i)}, \quad j = 1, \ldots, k, \quad t = 1, \ldots, n. \quad (3)$$

- Maximization step:

$$\hat{\alpha}_j \leftarrow \frac{1}{n} \sum_{t=1}^{n} w_{tj}, \qquad \hat{\mu}_j \leftarrow \frac{\sum_{t=1}^{n} w_{tj} x_t}{\sum_{t=1}^{n} w_{tj}}, \quad (4)$$

$$\hat{\Sigma}_j \leftarrow \frac{\sum_{t=1}^{n} w_{tj}(x_t - \hat{\mu}_j)(x_t - \hat{\mu}_j)^{\mathrm{T}}}{\sum_{t=1}^{n} w_{tj}}. \quad (5)$$

The first step in applying the EM algorithm to the problem at hand is to initialize the mixture model parameters. The K-Means algorithm is utilized to extract the data-driven initialization. The update scheme defined above allows for full covariance matrices; variants include restricting the covariance to be diagonal or scalar matrix. The updating process is repeated until the log-likelihood is increased by less than a predefined threshold from one iteration to the next. In this work, we choose to converge based on the log-likelihood and we use a 1% threshold. Other possibilities include converging based on the final probabilities extracted and using more strict convergence thresholds. We have found experimentally that the above convergence methodology works well for our purposes. Using EM, the parameters representing the Gaussian mixture are found. K-Means and EM are calculated for $k \geqslant 1$, with k corresponding to the model size.

We next wish to graphically illustrate the learned model in the color space. The parameter set that was estimated from the face-color training set is $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$. Denote

$$f_j(x \mid \alpha_j, \mu_j, \Sigma_j) = \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \times \exp\left\{ -\frac{1}{2}(x-\mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x-\mu_j) \right\}. \quad (6)$$

Eq. (6) provides a probabilistic representation for the affiliation of each input sample, $x = (r, g)$, to the Gaussian components that comprise the learned model. The probabilistic representation is complete, in that no information is lost.

It is often desired to proceed with a decision phase that is based on the extracted probabilities and provides a 'hard-decision' map of pixel affiliations into the predefined categories (in this case, face and no-face categories). Following the parameter estimation, we divide the color space into $k+1$ *decision regions*, k face-color regions and the complementary non-face region. The mapping into decision regions is based on the following steps:

- For each pair of $x = (r, g)$ coordinates in the color space, compute f_j, $j = 1, \ldots, k$.
- Find the Gaussian with the highest probability, l:

$$l = \arg\max_j f_j(x \mid \alpha_j, \mu_j, \Sigma_j). \tag{7}$$

- Mapping decision is based on the following rule:

$$\text{label}(x) = \begin{cases} l & \text{if } (x-\mu_l)^{\mathrm{T}} \Sigma_l^{-1}(x-\mu_l) < \text{thresh} \\ \text{non-face} & \text{else.} \end{cases} \tag{8}$$

For a probability that is less than a pre-determined, empirically defined threshold, the color-space coordinate is identified as a non-face-color coordinate. Otherwise, the color-space coordinate is labeled as a member of the selected Gaussian, l. In this work, we use a threshold of a single standard deviation, σ, throughout. This threshold is an empirical one. It may be possible to use a more data-driven threshold as well. In cases in which the face-color distribution is distinct from the background distribution (such as cases in which the background is well-controlled environment, e.g., a bluish screen in a portrait setting), it may be possible to increase the threshold (e.g., to 3σ), in which case we increase the probability of detection with no added risk of false alarms. In cases in which an overlap between the face color and background distributions is expected, a more conservative threshold should be used. A trade-off exists between the detection probability and the probability of false alarms. In this work, we use the more conservative threshold.

2.2. Experimental results in color modeling

In this section, we introduce the datasets that we are working with and present results of the database modeling via representative Gaussian mixtures.

A variety of databases were collected and experimented with. Two were downloaded from the website: http://rvl1.ecn.purdue.edu/ãleix/aleix_face_DB.html. The first set of people is normally lighted (ambient light) (designated 'ARH', which stands for Aleix, Robert, Homogeneous) while the second is with people lighted from left to right (designated 'ARL', which stands for Aleix, Robert, Light). An additional database (Video) is a batch of images taken from TV shows and movies. The last database (Mix) is composed as a set of images from both ARL and Essex. The Essex database was downloaded from the Essex University Vision Group at http://cswww.essex.ac.uk. The two databases are combined in order to synthetically generate a database of extreme colorings, with the ARL images being more yellowish and the faces in the Essex images having a more pinkish complexion. In Fig. 2, we show two images from the MIX database, demonstrating its variability. Table 1 presents a summary of the number of images and pixels in each of the above-described databases. Note that the model extraction (learning) is performed on a set of training images and a second, independent set of images in each database, serves as a testing set.

Fig. 2. Examples of faces present in the MIX database.

Table 1
Face databases

Database	# training images	# test images	# skin pixels
ARH	51	51	382K
ARL	42	128	267K
Video	82	82	8.2M
MIX	90	156	712K

We focus on the ARL database, in which strong lighting variations are present, and the MIX database, in which we artificially generated a database of extreme skin color variations. Decision maps for the ARL database are shown in Fig. 3 for $k = 1 - 3$, top down, respectively. On the left are

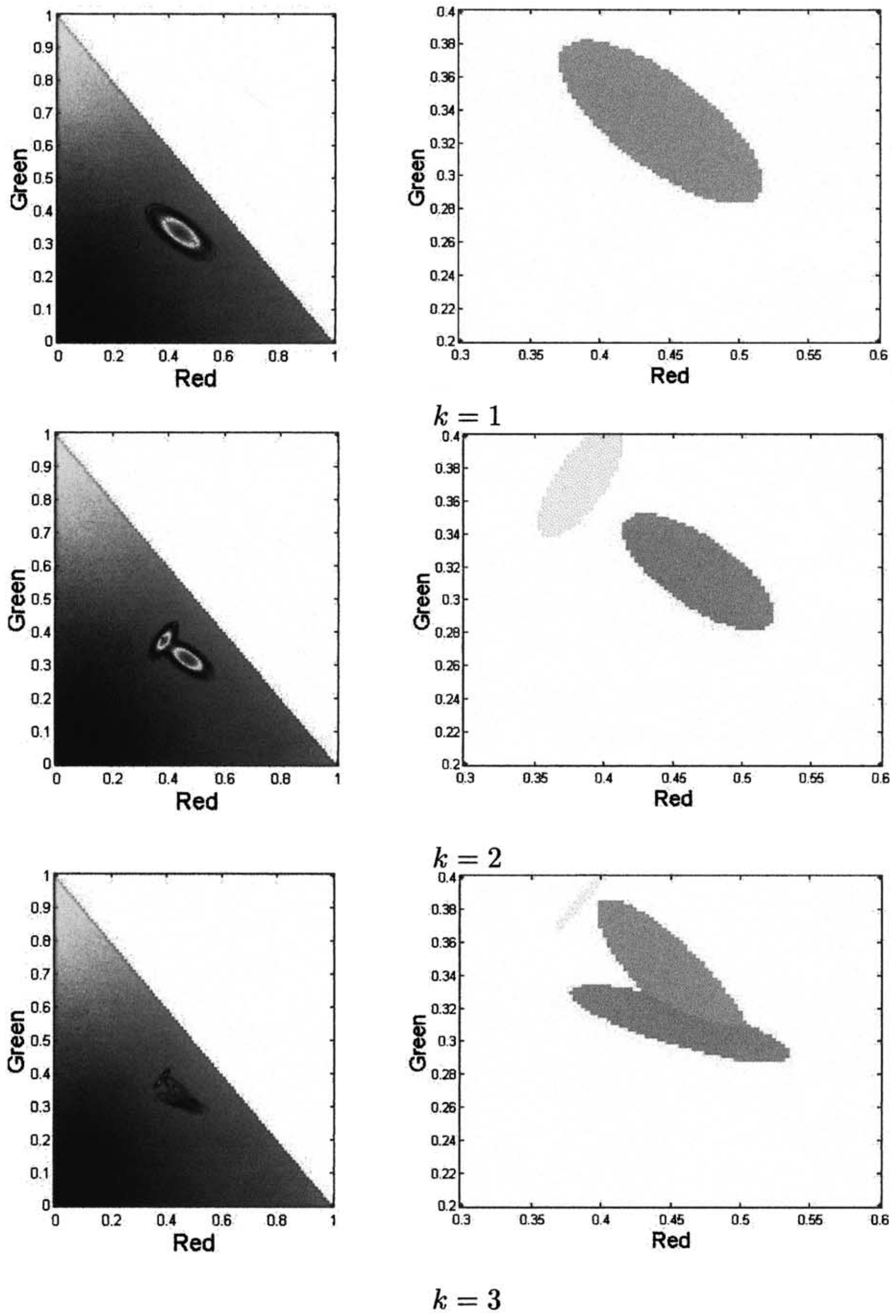

Fig. 3. Decision maps for differing values of k: $k = 1, 2, 3$ top bottom, respectively; ARL database used. On the left are the contour maps of the model for the corresponding k value, $\sigma = 1$. On the right are the support maps of the Gaussians in the model, shown with the respective mean color and size; a zoomed in view.

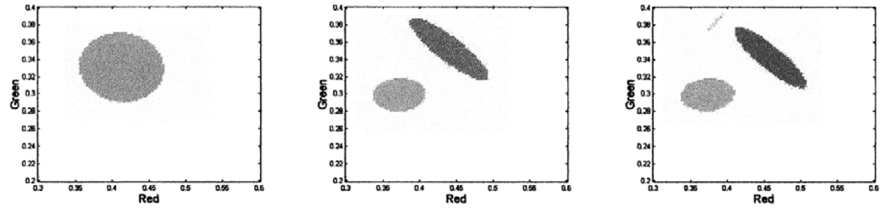

Fig. 4. Decision (support) maps for the MIX database, for $k = 1, 2, 3$ left to right, respectively.

the contour maps of the model for the corresponding k value, displayed in the color space. On the right, are the support maps of the Gaussians in the model, shown with the respective mean color and size. We note that for $k = 1$, the single Gaussian has a mean that represents the pinkish color, and a large overall support. In the $k = 2$ case, the support of the first Gaussian decreases as a second Gaussian models the more yellowish skin colors. The model changes further in the $k = 3$ case. Example of modeling the MIX database is shown in Fig. 4. The two main color distributions are extracted, with a yellowish source (the representative color in the ARL database) and a more pinkish source (representative color of the Essex database).

3. Face-color segmentation

An important benefit of the proposed approach is the ability to segment face regions with robustness to varying illumination conditions. Using the Gaussian mixture representation we are moreover able to segment *within-face* regions. We may automatically extract information that the left side of face is shadowed, the right part is highlighted and the face is of yellowish color. Such information may prove useful for a variety of follow-up image analysis objectives.

The labeling of individual pixels in the input image is accomplished via similar steps as used in the color space labeling process [Eqs. (7) and (8)]. Each pixel, $x = (r, g)$, in the image plane, is affiliated with the most probable Gaussian (closest in color space). If the probability of affiliation is below a decision threshold, the pixel is labeled as a non-face pixel.

A smoothing process follows the pixel-based classification, to provide contiguous label image regions. We use local neighborhood filtering (3×3 majority vote) and morphological operations of filling holes and removing isolated pixels to gen-

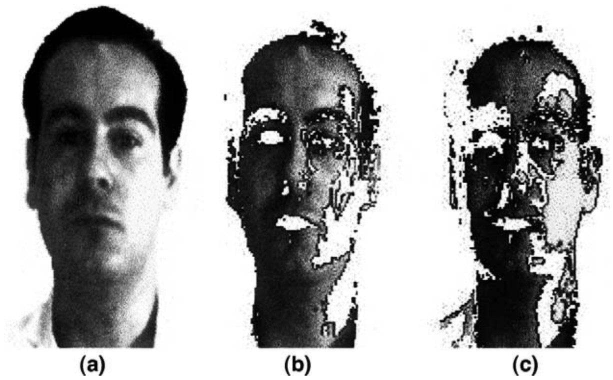

Fig. 5. Robustness to lighting variations. (a) The right side of the input face is illuminated while the left side remains with normal lighting conditions. (b) Single Gaussian modeling is used. With a unimodal Gaussian only the normal face half is extracted; half of the face is found to be a non-face region (shown as white pixels). (c) With Gaussian mixture modeling, the left half of the face is found to be associated with one Gaussian, while the right part of the face is found to be associated with a second Gaussian of the multi-modal mixture model. Each Gaussian is displayed with its mean color.

erate smooth regions in the final segmented image. More advanced smoothing methodologies may be used, including Markov-random fields (MRF) and Gibbs random field (GRF) model-based smoothing, in which the a posteriori probability of the segmentation labels is maximized, given the observed data along with spatial continuity constraints (Marchand-Maillet and Merialdo, 1999; Saber and Tekalp, 1998). We find the morphological filtering suffices to demonstrate the concepts we are presenting in this work.

The following examples demonstrate the possibility of achieving meaningful segmentation results that relate to lighting conditions. An example from the ARL database is presented in Fig. 5. The right side of the face is illuminated while the left side remains with normal lighting conditions (left image). This introduces a problem when a single Gaussian is used (center image). With a unimodal Gaussian only the normal face half is extracted; half of the face is found to be a non-face region (shown as white pixels). Using the Gaussian mixture model, the left half of the face is found to be affiliated with one Gaussian, while the right part of the face is found to be affiliated with a second Gaussian of the multi-modal mixture model (right image). In the pseudo-color result presented, each face-color pixel is displayed as belonging to one of the two Gaussians in the mixture. Each Gaussian is displayed with its mean color (the Gaussian that captures the more normal lighting face color is shown as a pinkish pseudo-color and the Gaussian that captures the more highlighted regions is displayed with a yellowish pseudo-color). A substantial part of the face is detected as a face region, which is the desired face segmentation result. Semantic meaning may be associated with the within-face segmented regions, according to their probabilistic affiliation with the varying Gaussian supports.

Additional examples of image segmentation into face-color regions are shown in Fig. 6. We see typical problematic cases with the single Gaussian case that are solved with the proposed mixture model. In Fig. 6(a) we see an instance from the ARL database in which highlights are present on both sides of the face. The highlighted regions are not extracted as face-color regions with the unimodal Gaussian model (shown as white pixels, center row). Thus, the remaining pixels that are extracted as face-color pixels do not form an oval shape. This may cause the face region as a whole

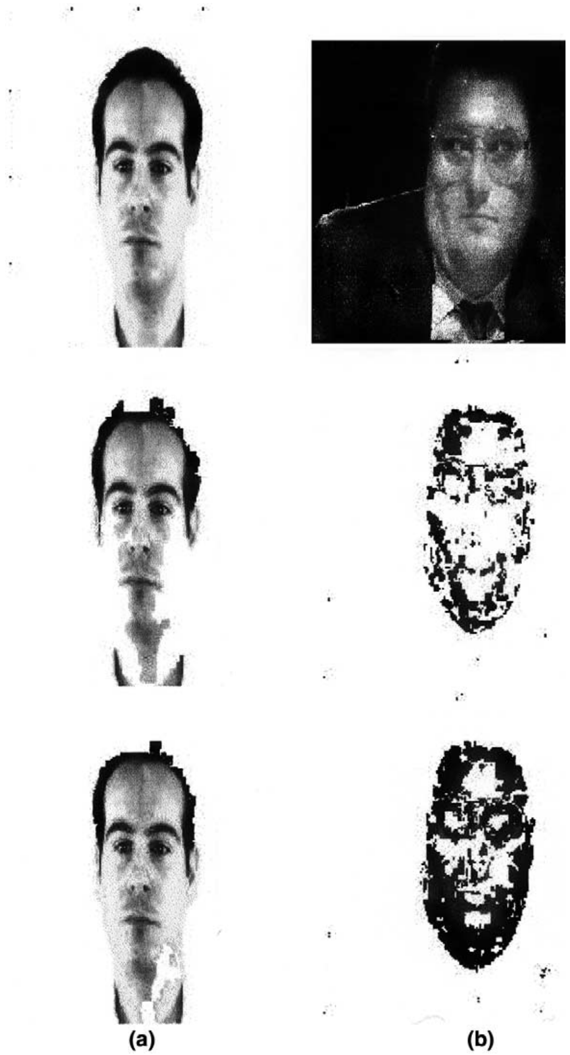

Fig. 6. Segmentation examples. In (a), we see an instance from the ARL database in which highlights are present on both sides of the face. The highlighted regions are not extracted as face-color regions with the unimodal Gaussian model (shown as white pixels, center row). Using the proposed modeling scheme, the entire face region is extracted (bottom row), as desired. In (b), we see an image from the Video database. Here many shadows exist. Using the unimodal Gaussian model (center row) we see a large portion of the pixels detected as non-face-color pixels (in white). This problem is solved using the presented model (bottom row).

not to be identified as a face in later processing. Using the proposed modeling scheme, the entire face region is extracted (bottom row), as desired. In Fig. 6(b), we see an image from the Video database. This image is a very challenging one in which many shadows exist across the face region. Using the unimodal Gaussian model (center row) we see a large portion of the pixels detected as non face-color pixels (in white), such that the overall number of extracted face-color pixels may not be sufficient for a face decision. The presented model provides a better solution to this very difficult task (bottom row).

Face detection examples with more complex backgrounds are presented in Fig. 7. In Fig. 7(I), we see a comparison between unimodal Gaussian modeling, Fig. 7(I)(b), and Gaussian mixture modeling, Fig. 7(I)(c). An additional face segmentation result using a Gaussian mixture model is shown in example Fig. 7(II)b. In all these examples skin-colored regions are extracted with minimal false detects of similar-color background regions.

4. Face archive statistical analysis

A second objective of this work is to show that the probabilistic Gaussian mixture model provides information about the variability present in the color space and the corresponding variability in a given input archive. A differentiation of interest may be the answer to the question: "What is a more important differentiating characteristic in the given archive: variations in color or variations in lighting?".

A statistical analysis of the pixel membership distribution per face, across all the images of the archive, enables to draw global conclusions about the dominant variability within a given archive. Each test image is segmented via the extracted model. For each face region we compute the

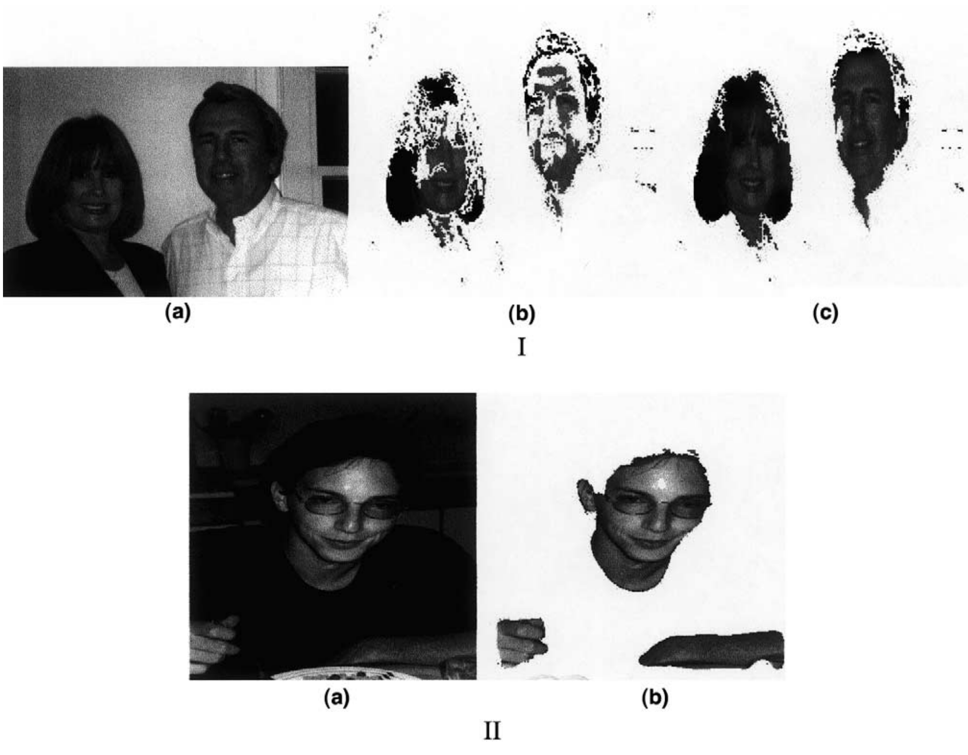

Fig. 7. More complex background examples. (I): (a) input image; (b) unimodal Gaussian model; (c) Gaussian mixture modeling. (II): (a) input image; (b) Gaussian mixture model output.

relative number of pixels affiliated with each Gaussian in the model. In the $k = 2$ scenario, e.g., for each image we extract the ratio number of pixels that have a color within the support of Gaussian number 1 to the number of pixels with a color belonging to the support of Gaussian number 2.

Fig. 8 presents a set of analysis graphs, for two different databases, the MIX database and the ARH database ($k = 2$). In Figs. 8(a) and (b), each sample point indicates the percentage of a face region that is affiliated to Gaussian number 1 (x-axis) and the percentage of the face region that is affiliated with Gaussian number 2 (y-axis). The ratio between the number of pixels that are associated with Gaussian number 1, to the total number of pixels in a single face is shown in the histograms of Figs. 8(c) and (d).

In Figs. 8(a) and (c) we see the results for the MIX database. The samples are seen to cluster in two main clusters, at the two extremes of the plot (0 and 1 in the histogram). Approximately, half of the face samples are affiliated with Gaussian number 1 and half are affiliated with Gaussian number 2. Each face region is uniformly associated with a single Gaussian in the mixture model, with no in-face segmentation. The distribution of the face samples matches with the fact that we have a mix of two distinct databases.

Figs. 8(b) and (d) display the results for the more homogeneous database, ARH. In this case, the samples are clustered in a single cluster in the center of the plot Fig. 8(b), and are spread around the 0.5 bin in the corresponding histogram. In this case, all the faces exhibit in-face splitting between the two Gaussians in the mixture.

The above results demonstrate that the more homogeneous archives, in which the variability within a face is greater than the variability across faces (such as the case with lighting variations), will exhibit splitting of the face regions, resulting in support for multiple Gaussians in the model. Alternatively, archives that have strong variations in color space (e.g., combine several ethnic groups), such that the variability between faces is more dominant than the in-face variability, will tend to result in face regions that are mostly homogeneous in the pixel memberships to one

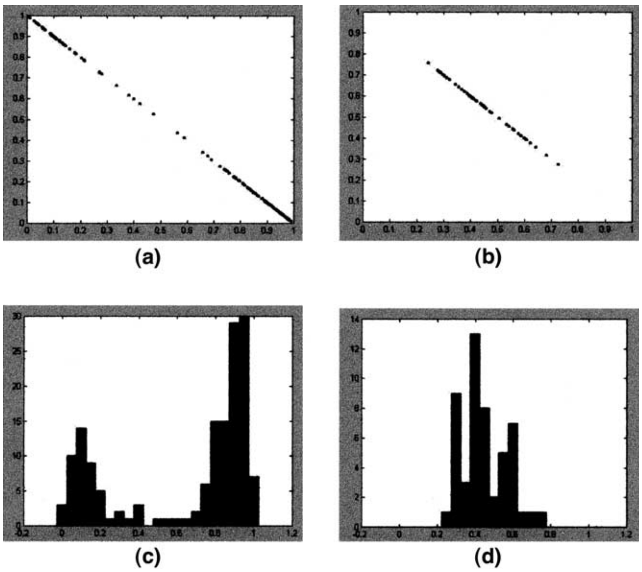

Fig. 8. Statistical analysis for the MIX database ((a), (c)) and the ARH database ((b), (d)). In (a) and (b), each sample point indicates the percentage of a face region that is affiliated with Gaussian number 1 (x-axis) and the percentage of the face region that is affiliated with Gaussian number 2 (y-axis). The ratio between the number of pixels that are associated with Gaussian number 1, to the total number of pixels in a single face is shown in the histograms (c) and (d).

particular Gaussian (the Gaussian in the model that best fits the persons face color).

5. Discussion

We present a general methodology for color modeling within a given archive, that allows for face-color modeling and segmentation. The proposed shift from a unimodal p.d.f. to a multimodal p.d.f., allows for better modeling and augments analysis capabilities.

In this work, we demonstrated that via the multi-mode semi-parametric modeling, we have sufficient descriptive power to capture multiple variations within the given image archive, such as variations of natural color, as well as variations in the lighting conditions. We have demonstrated the ability to provide better overall face segmentation results, as well as the ability to analyze the different regions within a face as belonging to different color and lighting categories. These characteristics are helpful in further higher-level processing, such as facial feature extraction (e.g., eyes, nose, mouth), and complete face detection systems. We have shown a statistical analysis of a set of images passing thru the model and have demonstrated that global conclusions about the dominant variability within the archive can be extracted.

In this work, we focused on the case of $k = 2$. We intend to explore the added value in increasing the number of components further. We are also investigating means of automatically determining the most appropriate number of model components per given archive. The probabilistic framework used in this work gives a probabilistic affiliation of the image pixels to the learned model Gaussians. In segmentation systems in which a hard-decision (face, no-face) is desired, thresholds need to be utilized. In this work, we use empirically defined thresholds. A shift to more data-driven thresholds is desired. The proposed modeling scheme is not limited to face content. The methodology presented is a general one and can be extended to additional image archive modeling. Future work entails testing on larger face archives, with a large number of variability sources.

References

Burl, M.C., Leung, T.K., Perona, P., 1995. Face localization via shape statistics. In: Internat. Workshop on Automatic Face and Gesture Recognition, June 1995.

Chang, T.C., Huang, T.S., Novak, C., 1994. Facial feature extraction from color images. In: Proc. Internat. Conf. on Pattern recognition (ICPR'94), Jerusalem, Israel, Vol. B., pp. 39–43.

Dai, Y., Nakano, Y., 1995. Extraction of facial images from comlex background using color information and sgld matrices. In: Proc. Internat. Workshop on Automatic Face- and Gesture-recognition, Zurich, pp. 238–242.

Dempster, A., Laird, N., Rubin, D., 1997. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39 (1), 1–38.

Garcia, C., Tziritas, G., 1999. Face detection using quantized skin color regions merging and wavelet packet analysis. In: Conf. on Multimedia Computing and Systems, Vol. I, pp. 703–708.

Jeng, S.H., Yao, H.Y.M., Han, C.C., Chern, M.Y., Liu, Y.T., 1998. Facial feature detection using geometrical face model: an efficient approach. Pattern Recognition 31 (3), 273–282.

Kim, H.S., Kim, N.K., Ahn, S.C., Kim, H.G., 1998. Object oriented face detection using range and color information. In: Proc. Third Internat. Workshop on Automatic Face- and Gesture-recognition, Japan, pp. 112–117.

Lee, C.H., Kim, J.S., Park, K.H., 1996. Automatic face location in a complex background using motion and color information. Pattern Recognition 29 (11), 1877–1889.

Lin, S.H., Kung, S.Y., Lin, L.J., 1997. Face recognition/detection by probabilistic decision-based neural-network. IEEE Trans. Neural Networks 8 (1), 114–131.

Marchand-Maillet, S., Merialdo, B., 1999. Pseudo two-dimensional hidden Markov models for face detection in color images. In: Second Internat. Conf. on Audio- and Video-based Biometric Person Authentication, pp. 13–18.

McKenna, S., Gong, S., Raja, Y., 1998. Modelling facial color and identity with Gaussian mixtures. Pattern Recognition 31 (12), 1883–1892.

Phillips, P.J., McCabe, R., Chellappa, R., 1998. Biometric image processing and recognition. In: Proc. IX European Signal Processing Conference, I.

Rowly, H.A., Baluja, S., Kanade, T., 1998. Neural-network based face detection. IEEE Trans. Pattern Anal. Machine Intell. 20, 23–28.

Saber, E., Tekalp, A.M., Eschbach, R., Knox, K., 1996. Automatic image annotation using adaptive color classification. Graph. Models Image Process. 58 (2), 115–126.

Saber, E., Tekalp, A.M., 1998. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. Pattern Recognition Lett. 19, 669–680.

Sigal, L., Sclaroff, S., Athitsos, V., 2000. Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2000), SC.

Sobottka, K., Pitas, I., 1996. Segmentation and tracking of faces in color images. In: Proc. Second Internat. Workshop on Automatic Face- and Gesture-recognition, Zurich, pp. 344–349.

Sun, Q.B., Huang, W.M., Wu, J.K., 1998. Face detection based on color and symmetry information. In: Proc. Third Internat. Workshop on Automatic Face- and Gesture-recognition, Japan, pp. 130–135.

Sung, K.K., Poggio, T., 1998. Example-based learning for view-based human face detection. IEEE Trans. Pattern Anal. Machine Intell. 20, 39–51.

Terrillon, J.C., Akamatsu, S., 1998. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In: Proc. Third Internat. Workshop on Automatic Face- and Gesture-recognition, Japan, pp. 112–117.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cognitive Sci. 3 (1), 71–86.

Wang, J.G., Sung, E., 1999. Frontal-view face detection and facial feature extraction using color and morphological operation. Pattern Recognition Lett. 20 (10), 1053–1068.

Wei, G., Sethi, I.K., 1999. Face detection for image annotation. Pattern Recognition Lett. 20, 1313–1321.

Wiskott, L., Fellous, N., Kruger, N., Von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 775–779.

Wu, H., Chen, Q., Yachida, M., 1999. Face detection from color images using a fuzzy pattern matching method. IEEE Trans. Pattern Anal. Machine Intell. 21 (6).

Yang, G., Huang, T.S., 1994. Human face detection in a complex background. Pattern Recognition 27 (1), 53–63.

Yang, M-H., Ahuja, N., 1999. Gaussian mixture model for human skin color and its application in image and video databases. In: Proc. SPIE: Storage and Retrieval for Image and Video Databases (SPIE 99), San Jose, pp. 458–466.

Yang, J., Lu, W., Waibel, A., 1998. Skin-color modeling and adaptation. In: Proc. Third Internat. Workshop on Automatic Face- and Gesture-recognition, Japan.

Yow, K.C., Cipolla, C., 1997. Feature-based human face detection. Image Vision Comput. 15, 713–735.