EE 368 Project: Face Detection in Color Images

Wenmiao Lu   and   Shaohua Sun

Department of Electrical Engineering

Stanford University

May 26, 2003

Abstract

We present in this report an approach to automatic detection of human faces in color images. The proposed approach consists of three parts: a human skin segmentation to identify probable regions corresponding to human faces; adaptive shape analysis to separate isolated human faces from initial segmentation results; and a view-based face detection to further identify the location of each human face. The human skin segmentation employs a model-based approach to represent and differentiate the background colors and skin colors. To further refine the initial segmentation results, the adaptive shape analysis uses a series of morphological operations based on the prior shape knowledge of upright human faces. The view-based face detection is constructed based on principal component analysis and neural network classification. Our face detector has been applied to several test images, and satisfactory results have been obtained.

Index Terms – Human skin segmentation, Gaussian mixture model, Adaptive shape analysis, Principal component analysis, and Neural network.

I. INTRODUCTION

We present in this report an approach to automatic detection of human faces in color images. The proposed approach consists of three parts: a human skin segmentation to identify probable regions corresponding to human faces; adaptive shape analysis to identify isolated human faces in the initial segmentation results; and a view-based face detection to further identify the location of each human face. The human skin segmentation employs a model-based approach to represent and differentiate the background colors and skin colors. To further refine the initial segmentation results, the adaptive shape analysis uses a series of morphological operations based on the prior shape knowledge of upright human faces. The view-based face detection is constructed based on principal component analysis and neural network classification.

The remainder of the report is organized as follows. The human skin segmentation is given in Section II. Section III presents the adaptive shape analysis of initial binary segmentation map, and Section IV describes the face detection based on principal component analysis and neural network classification. We summarize our work and report experimental results in Section VI.

II. HUMAN SKIN SEGMENTATION

Color is a prominent feature of human faces. Using skin color as a primitive feature for detecting face regions has several advantages. In particular, processing color is much faster than processing other facial features. Furthermore, color information is invariant to face orientations. However, even under a fixed ambient lighting, people have different skin color appearance. In order to effectively exploit skin color for face detection, we need to find a feature space, in which human skin colors cluster tightly together and reside remotely to background colors.

We adopt the YC_bC_r color space since it is perceptually uniform and separates luminance and chrominance. Many research studies [1] found that the chrominance components of the skin-tone color are independent of the luminance component. Hence, in our implementation, only the C_bC_r components (the chrominance components) are used to model the distribution of skin colors.

In the C_bC_r subspace, the distribution of skin colors is modeled with a multivariate Gaussian mixture model (GMM), thereby the parameters of the Gaussian mixture model are estimated using the standard Expectation-Maximization (EM) algorithm [2]. That is, each skin color value is viewed as a realization from a Gaussian mixture model $\mathcal{S}$ consisting of Gaussian components $\{\mathcal{S}_i\}_{i=1}^{k}$, each characterized by its mean vector $\mu_{\mathcal{S}_i}$ and covariance matrix $\Sigma_{\mathcal{S}_i}$, in some proportions $\gamma_1, \ldots, \gamma_k$, where $\sum_{i=1}^{k} \gamma_i = 1$ and $\gamma_i > 0$; the number of Gaussian components k is 4 in our implementation. The probability that a pixel j with color value X_j belongs to the skin

1

color model $\mathcal{S}$ can be computed as

$$p(X_j|\mathcal{S}) = \sum_{i=1}^{k} \gamma_i p(X_j|\mathcal{S}_i) = \sum_{i=1}^{k} \frac{\gamma_i}{(2\pi)^{3/2}|\Sigma_{\mathcal{S}_i}|^{1/2}} \exp\left\{-\frac{1}{2}(X_j - \mu_{\mathcal{S}_i})^T \Sigma_{\mathcal{S}_i}^{-1}(X_j - \mu_{\mathcal{S}_i})\right\} \tag{1}$$

Similarly, the pixels of the background scene are also modeled with a GMM. To account for the variety of background colors, the number of Gaussian components in the background GMM is larger than that of the skin color. In our implementation, the number of Gaussian components is set to 8 for the background GMM. Let $\mathcal{B}$ denote the background model consisting of a mixture of Gaussian components $\{\mathcal{B}_i\}_{i=1}^{k}$, each characterized by its mean vector $\mu_{\mathcal{B}_i}$ and covariance matrix $\Sigma_{\mathcal{B}_i}$, in some proportions $\alpha_1, \ldots, \alpha_k$, where $\sum_{i=1}^{k} \alpha_i = 1$ and $\alpha_i > 0$. The probability that a pixel j with color value X_j belongs to the background model $\mathcal{B}$ can be computed as

$$p(X_j|\mathcal{B}) = \sum_{i=1}^{k} \alpha_i p(X_j|\mathcal{B}_i) = \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{3/2}|\Sigma_{\mathcal{B}_i}|^{1/2}} \exp\left\{-\frac{1}{2}(X_j - \mu_{\mathcal{B}_i})^T \Sigma_{\mathcal{B}_i}^{-1}(X_j - \mu_{\mathcal{B}_i})\right\} \tag{2}$$

After obtaining the GMMs of skin colors and background colors, the segmentation of human faces can then be done by maximum likelihood classification of pixels within a test image. Specifically, given the background model $\mathcal{B}$ and the skin model $\mathcal{S}$, a pixel with color value X is considered to be skin pixel if

$$p(X|\mathcal{S}) > p(X|\mathcal{B}) \tag{3}$$

Figure 1 shows an initial segmentation map of a test image produced by the proposed skin segmentation method. Subsequently, this initial binary segmentation map is subject to the following analysis by incorporating the prior shape knowledge of upright human faces.

Figure 1: Initial binary segmentation map of a test image produced by the proposed skin segmentation.

III. ADAPTIVE SHAPE ANALYSIS

In the stage of adaptive shape analysis, initial binary segmentation map is processed by a series of morphological operations to separate regions corresponding to isolated faces. Furthermore, this stage determines which regions should be passed to the next stage for final detection. To this end, we implement the following three steps:

(a) Erosion and dilation to refine the binary map [3];

(b) for each connected region in the binary map, use OPEN [3] operation to further divide them into smaller regions; and

(c) for the output map of the previous step (b), identify isolated regions with medium size and shape similar to that of a single face.

Those isolated regions with medium size and "face" shape are deemed as detected faces; hence, only those un-identified regions, which normally have odd sizes and shapes, are passed to next stage for final face detection.

Figure 2 shows the refined binary map of Figure 1 produced by the step (b). In this step, each single region in the initial binary map is processed with OPEN operation in both horizontal and vertical directions. As a result, with carefully chosen OPEN structuring elements, the connected regions that contains more than one faces can be separated into isolated faces. In Figure 2, every circle indicates isolated faces separated from previously connected regions.

Figure 3 shows the effect of the step (c). In this image the connected regions smaller than a certain size have been removed. Consequently, the regions left in the binary map are good face candidates. However we still cannot differentiate faces and non-faces, and face locations remain unknown. For those regions with large sizes, they may correspond to several connected faces, while regions with small sizes can be some body parts, for example, hands. To resolve such ambiguity, those regions undergo further analysis – a view-based face detection.

IV. VIEW-BASED FACE DETECTION

While the adaptive shape analysis of the initial binary segmentation map effectively identifies most isolated faces, the locations of connected faces remain unsolved. Figure 3 shows several connected faces. In such a case, the initial skin segmentation is not able to separate each individual face, and the shape analysis does not resolve the locations of constituent faces. To remedy this situation, the connected face regions are subject to the further analysis based on the principal component analysis (PCA) and neural network classification. As such analysis is based on frontal view of human face samples, it is also commonly referred to as view-based face detection.

This part of the work is largely the reproduction of the method proposed by Sung K. K. [4]. A 19×19 window slides through the input image to identify possible face locations. The face detection is done by transforming
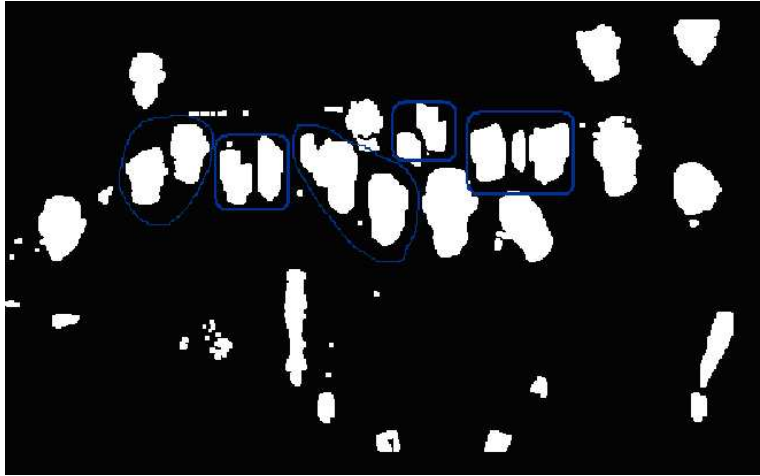
Figure 2: The refined binary map with isolated faces found by OPEN operation.

Figure 3: The refined binary map in which connected regions smaller than a certain size have been removed.

each window pattern into a low-dimension feature space and comparing the transformed window pattern with a canonical face model. Each 19×19 window pattern is treated as a point in the feature space, and the face detection task is to determine if a coming data point falls into the region corresponding to face patterns.

Hence, we need to infer the region corresponding to face patterns in the multi-dimensional (in this case $19 \times 19 = 361$) feature space in a tractable fashion. That is, the distribution of face patterns in the multidimensional feature space is to be represented by fitting training face samples with a face model. As the distribution of face patterns is rather complicated, the face model adopted consists of a few multi-dimensional Gaussian clusters (6 in our implementation). To further carve out the regions nearby the face region, we also use 6 "non-face" Gaussian clusters to refine the boundary of the face region. Each cluster is characterized by its mean vector and covariance

matrix. The mean vector can be viewed as the prototype of the associated cluster, while the covariance matrix encodes the pattern variation within the cluster.

Subsequently, we are to match each candidate window pattern against the face model. Each match captures the degree of "similarity" between the test pattern and the face model. As the face model consists of 12 clusters (6 face clusters and 6 non-face clusters), each matching metric is a vector of distances between the test pattern and the prototypes of the 12 clusters. With regarding to the distance measurements to one single cluster, it contains two values, namely, the normalized Mahalanobis distance and the projected Euclidean distance.

The first distance is the normalized Mahalanobis distance between the cluster prototype and a test pattern's projection in a subspace spanned by its principal eigenvectors. The normalized Mahalanobis distance between the column vector test pattern $\vec{X}$ and the prototype pattern $\vec{\mu}$ is given by

$$\mathcal{D}_1 = \sum_{i=1}^{60} (\ln \lambda_i + \frac{\|\psi_i^T (\vec{x} - \vec{\mu})\|^2}{\lambda_i}), \tag{4}$$

where λ_i and ψ_i are the ith largest eigenvalue and its associated eigenvector, respectively. Therefore, this distance is analogous to the $distance-in-feature-space$ in classical principal component analysis. In our implementation, due to limited training face samples, the subspace of each cluster is spanned by 60 largest eigenvectors.

The second distance is the projected Euclidean distance that measures the reconstruction quality of the test pattern using those largest eigenvectors. As compared to the first distance, this measures the $distance-from-feature-space$, which is given by

$$\mathcal{D}_2 = \|(I - E_{60} E_{60}^T)(\vec{x} - \vec{\mu})\|, \tag{5}$$

where the ith column of E_{60} is the ith principal eigenvector. Intuitively, a face pattern with minor spatial and intensity variations still resemble canonical face patterns, while a non-face pattern would less likely appear as a valid face pattern. Therefore, when reconstructed using the principal eigenvectors of a face cluster, a face pattern is likely to have much less reconstruction error as compared to a non-face pattern.

Finally, we need a classifier to distinguish face patterns from non-face patterns based on their matching distances to 12 clusters of our face model. This classification task is done by a Multi-layer neural network. During the classification, the network is given a vector of the test pattern's matching distances to 12 clusters. The output unit returns a '1' if the input distances arise from a "face" pattern, and a '-1' otherwise.

V. Experimental Results and Concluding Remarks

We present in this section the experimental results of the seven images. Figure 4 shows one sample detection result. This image contains human faces with different sizes and orientations, which imposes great challenge

on our face detector. Nevertheless, our face detector manages to detect all the faces without any false detection. For the other six images, our face detector also produces very satisfactory results, which are listed in Table 1. In summary we have presented a hybrid approach to automatic detection of human faces in color images and demonstrate the efficacy of our face detector by several test images.

Table 1: Overall Performance of the Face Detector Applied to Seven Images.

	Hits	Repeated	False	Missed	Time(seconds)
Training1	21	0	0	0	164.6
Training2	23	0	0	1	33.5
Training3	24	0	0	1	79.5
Training4	22	0	1	2	19.1
Training5	22	0	0	2	49.2
Training6	23	0	0	1	47.3
Training7	22	0	0	0	149.7

Figure 4: Sample Detection Result.

REFERENCES

[1] J. Yang, W. Lu, and A. Waibel, "Skin-Color Modeling and Adaptation", *ACCV*, pp. 687-694, 1998.

[2] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. Wiley Interscience, 1997.

[3] A. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall Press, 1989.

[4] K.K. Sung and T. Poggie, "Example-based Learning for View-based Human Face Detection ," *IEEE Trans. on PAMI*, vol. 20, no. 1, 1998.