Detecting Human Faces in Color Images

Ming-Hsuan Yang and Narendra Ahuja
Beckman Institute and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801
E-mail: {mhyang, ahuja}@stereo.ai.uiuc.edu

Abstract

We propose a new method to detect human faces in color images. A human skin color model is built to capture the chromatic properties based on multivariate statistical analysis. Given a color image, multiscale segmentation is used to generate homogeneous regions at multiple different scales. From the coarsest to the finest scale, regions of skin color are merged until the shape is approximately elliptic. Postprocessing is performed to determine whether a merged region contains a human face and include the facial features of non-skin color such as eyes and mouth if necessary. Experimental results show that human faces in color images can be detected regardless of size, orientation and viewpoint.

1 Introduction

Face detection has many applications, including teleconferencing [2], face recognition [6], and gesture recognition [12]. The goal of face detection is to determine whether or not there is any human face in the image, and, if present, return its location and spatial extent. The task involves scale, space, orientation, and viewpoint analysis since faces of different sizes may appear in arbitrary locations and orientations in an image. In recent years, many methods have been proposed to detect human faces in a single image [6, 8, 4] or a sequence of images [11] based on gray scale [6, 8, 4] or color [2, 11]. Yang and Huang [9] develop a rule-based system to detect faces in that face candidates are determined using coded rules and these candidates are classified as face or non-face using edge-based features. Sung and Poggio [8] report an example-based learning approach for locating vertical frontal view of human faces. They model the distribution of human face patterns by means of a few view-based face and non-face prototype clusters. A small window is moved over all portions of an image and determines whether a face exists in each window based on distance metrics. Similar approach is adopted and extended by Rowley, Baluja and Kanade [6]. Support vector machine has also been applied to face detection [4]. Yang and Waibel [11] use a pixel-based operation to cluster skin-like pixels for face tracking in that motion and geometric information is also employed. Yow and Cipolla [13] present a feature-based algorithm for detecting faces in which feature points are detected using spatial filters and then grouped into face candidates using geometric and gray level constraints. A probabilistic framework is then used to reinforce probabilities and to evaluate the likelihood of the candidate as a face. Most recently, several modular systems using a combination of shape analysis, color segmentation and motion information for locating or tracking heads and faces in an image sequence have been developed [2, 12, 11, 5, 7].

Most of the existing systems use window-based or pixel-based operations to detect faces. The problem with window-based approach is that it cannot detect faces of different orientations or view angles. On the other hand, pixel-based operation is ineffective in distinguishing a human face from other skin areas such as hands and arms in an image. We propose a new method to detect human faces to deal with these problems using multiscale segmentation, color, and geometric information. Multiscale segmentation is utilized to generate homogeneous regions that can deal with the scale problem. Under white lighting conditions, color does not vary significantly with orientation or view angles. Thus, we build a skin color model to capture the chromatic characteristics using multivariate statistical analysis. The analysis shows that human skin colors fall in a small region in color space and can be approximated by a Gaussian distribution. Thus, color features of human faces are employed to cope with problems of orientation and view angles. Given a color image, a human faces is detected if the shape of the merged skin color regions can be approximated by an ellipse and there exist some darker fa-

0-8186-8821-1/98 $10.00 © 1998 IEEE

cial features, such as eye bows and eyes, inside that region. Our experimental results show that human faces in color images can be identified regardless of size, orientation and viewpoint.

2 Proposed Method

2.1 Multiscale Segmentation

Multiscale image segmentation is used to extract a hierarchy of regions for matching. Segmentation is achieved using a transform function, which maps the feature primitives to a family of attraction force fields, defined by

$$\mathbf{F}(x, y; \sigma_g(x,y), \sigma_s(x,y)) = \int\int_R d_g(\Delta I, \sigma_g(x,y)) \cdot d_s(\vec{r}, \sigma_s(x,y)) \frac{\vec{r}}{||\vec{r}||} dw dv \tag{1}$$

where $R = domain(I(u,v)) \backslash \{(x,y)\}$ and $\vec{r} = (v - x)\vec{i} + (w - y)\vec{j}$. The details of this transform and the segmentation method can be found in [1]. Here we review its basic characteristics. The parameter σ_g is a homogeneity scale which reflects the homogeneity of the region into which pixel groups and σ_s is a spatial scale that controls the neighborhood from which the force on the pixel is computed. The force field encodes the region structure in a manner which allows easy extraction.

The homogeneity between two pixels is given by the Euclidean distance between the associated m-dimensional feature vectors (e.g., $m = 3$ for a color image)

$$\Delta I = |I(x,y) - I(v,w)| \tag{2}$$

The spatial scale parameter, σ_s, controls the spatial distance function, $d_s(\cdot)$, and the homogeneity scale parameter, σ_g, controls the homogeneity distance function, $d_g(\cdot)$. One possible form for these functions satisfying criteria discussed in [1] are unnormalized Gaussian

$$\begin{aligned} d_g(\Delta I, \sigma_g) &\sim \sqrt{2\pi\sigma_g^2} N_{\Delta_I}(0, \sigma_g^2) \\ d_s(\vec{r}, \sigma_s) &\sim \begin{cases} \sqrt{2\pi\sigma_s^2} N_{||\vec{r}||}(0, \sigma_s^2), & ||\vec{r}|| \leq 2\sigma_s \\ 0, & ||\vec{r}|| > 2\sigma_s \end{cases} \end{aligned} \tag{3}$$

With the above definition of force field $\mathbf{F}$, pixels are grouped together into regions whose boundaries correspond to diverging force vectors in $\mathbf{F}$ and whose skeletons correspond to converging force vectors in $\mathbf{F}$. An increase in σ_g causes less homogeneous structures to be encoded and an increase in σ_s causes large structures to be encoded.

Figure 1 (a) shows an original image. The following three images show the segmented image with all pixels in regions at three different scales ($\sigma_g = 40, 20, 6$ for (b), (c), and (d)) replaced by a unique gray value for each region. The identified regions are visually correct and the region boundaries align with the actual boundaries at all scales. Note that all segmentation parameters of the transform are selected automatically, eliminating the need to make *a priori* assumptions about either the geometric or homogeneity characteristics of the structure.

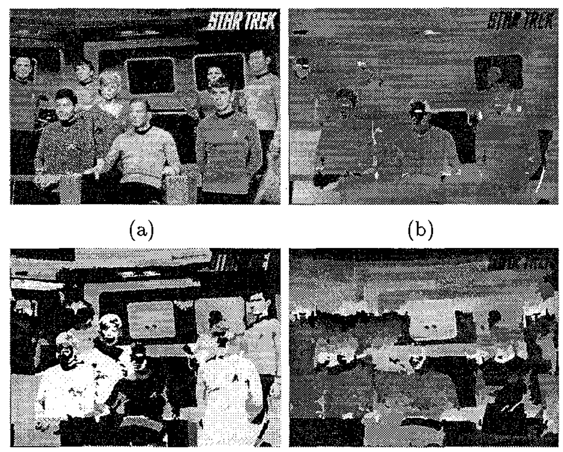


Figure 1: Results of multiscale segmentation

2.2 Human Skin Color Model

Although different people have different colors in appearance, several studies have showed the major difference lies in intensity rather than color itself [2, 10]. To build a skin color model, we use CIE LUV color space and discard the luminance value. Based on the histogram of more than five hundred images that contain human skins of different races, we observe that the distribution of skin color, $\mathbf{x} = (\mathbf{u}, \mathbf{v})^{\mathbf{T}}$, can be modeled by a Gaussian distribution. Figure 2 shows a uniformly subsampled color. Therefore, we hypothesize the distribution of skin color as a bivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ where $\mu = (\mu_\mathbf{u}, \mu_\mathbf{v})^{\mathbf{T}}$ and $\Sigma = \begin{vmatrix} \sigma_{uu}^2 & \sigma_{uv}^2 \\ \sigma_{vu}^2 & \sigma_{vv}^2 \end{vmatrix}$. The null hypothesis of Gaussian distribution is accepted using a chi-square test. A recent report [10] on skin color modeling using RGB color space has similar results. The parameters of the distribution are obtained by maximum likelihood estimation. A pixel is identified to have skin color if its corresponding probability is greater than a threshold (0.5) and a region is classified to have human skin color if most (above 70%) of its pixels have skin color. Figure 3(a) shows the segmented regions of Figure 1(a) that have skin color. Note that skin color helps in detecting human face, but skin color alone cannot detect human faces correctly.

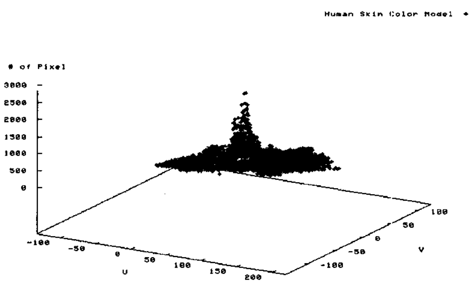


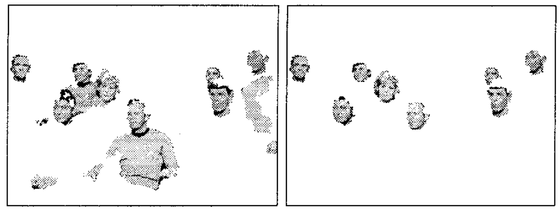
Figure 2: Distribution of human skin color

2.3 Geometric Analysis and Postprocessing

Since the shapes of human faces are approximately elliptic, skin color regions are merged until the shape of the merged region is approximately elliptic. The orientation of an ellipse can be calculated by the moments of inertia [3]. The extents of the major and minor axes of the ellipse are approximated by the extents of the region along the axis directions and the degree of fit of the ellipse is determined by the number of pixels that fall into that shape specified by the computed parameters. A merged skin color region is a human face candidate if the ratio of major axis to minor axis is less than a threshold (1.7). If a candidate region has some darken regions or some holes inside the merged region, it is classified as a human face. The rationale is that a face contains eyes or mouth and these facial features have darker colors than skin. If these facial features are included in the skin color region, the color of the facial features should be relatively darker than the rest of the region. If the facial features are not classified as skin color regions, some empty areas should exist inside the candidate region. In this case, the segmented regions of the empty areas are included as part of a face. Figure 3(b) shows the final result of detected human faces in Figure 3.

3 Experimental Results

In this section, we present experimental results (color images of the experiments can be found at http://uirvli.ai.uiuc.edu/mhyang/facedetection.html)



(a) (b)

Figure 3: Intermediate and final results of face detection

on color images collected from the Internet. Each detected human faces is shown with an enclosing ellipse. Figure 4(a) shows that human faces of different viewpoints and sizes are detected. Note that faces from different ethnic backgrounds are detected. Faces of different orientations can also be identified as shown in Figure 4(b) Figure 5 gives an example that human faces with glasses and sideburns can still be detected. Finally, Figure 6 demonstrates that human faces can be extracted in a color image with complex background.

(a) (b)

Figure 4: Faces of different size and orientation

4 Conclusion

In this paper, we have presented a new method to detect human faces in color images. The proposed method utilizes a multiscale transform to segment images into homogeneous regions and extracts skin regions based on a skin color model. These regions are then merged until the shape is approximately elliptic and verified. Experimental results show that the proposed method can detect human faces in color image regardless of size, orientation and viewpoint.

Acknowledgments

The support of the Office of Naval Research under grant N00014-96-1-0502 is gratefully acknowledged.

Figure 5: Faces with different features

Figure 6: Human faces in complex background

References

[1] N. Ahuja. A transform for multiscale image segmentation by integrated edge and region detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(12):1211–1235, December 1996.

[2] H. P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan. Multimodal system for locating heads and faces. In *Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition*, pages 88–93, 1996.

[3] B. Horn. *Robot vision*. McGraw-Hill, 1986.

[4] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136, 1997.

[5] Y. Raja, S. McKenna, and S. Gong. Segmentation and tracking using colour mixture models. In *Proceedings of the third Asian Conference on Computer Vision*, pages 607–614, 1998.

[6] H. Rowley, S. Baluja, and T. Kanade. Human face detection in visual scenes. In *Advances in Neural Information Processing Systems*, volume 8, 1997.

[7] J. Sobottka and I. Pittas. Segmentation and tracking of faces in color images. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pages 236–241, 1996.

[8] K.-K. Sung and T. Poggio. Example-based learning for view based human face detection. Technical Report AI Memo 1521, MIT, 1994.

[9] G. Yang and T. S. Huang. Human face detection in complex background. *Pattern Recognition*, 27(1):53–63, 1994.

[10] J. Yang, W. Lu, and A. Waibel. Skin-color modeling and adaptation. Technical Report CMU-CS-97-146, School of Computer Science, Carnegie Mellon University, 1997.

[11] J. Yang and A. Waibel. A real-time face tracker. In *Proceedings of the Third Workshop on Applications of Computer Vision*, pages 142–147, 1996.

[12] M.-H. Yang and N. Ahuja. Extracting gestural motion trajectory. In *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*, pages 10–15, 1998.

[13] K. C. Yow and R. Cipolla. Feature-based human face detection. *Image and Vision Computing*, 15(9):713–735, 1997.