Color-Based Face Detection using Skin Locus Model and Hierarchical Filtering

A. Hadid, M. Pietikäinen, and B. Martinkauppi
Machine Vision Group
Infotech Oulu, University of Oulu, P.O.Box 4500, FIN-90014, FINLAND
{hadid,mkp,jbm}@ee.oulu.fi

Abstract

This paper introduces a new architecture for face detection in color images. Based on skin locus and successive detectors, the method allows high efficiency under drastically varying illumination conditions. While most color-based approaches work only if the camera was calibrated and detect faces if the features (eyes, mouth, eyebrows…) are visible and detectable, our system detects faces even when these features are not well defined. The detection is based on a robust modeling of skin color, called skin locus, which is used to extract the skin-like regions. After orientation normalization and based on verifying a set of criteria (face symmetry, presence of some facial features, variance of pixel intensities and connected component arrangement), only facial regions are selected. We show that our system can detect faces and deal with different conditions (size, orientation, illumination, and complex background).

1. Introduction

Accomplishing face detection from single images is a challenging task because of variability in scale, orientation, pose, facial expressions, lighting conditions, and camera calibrations. In recent years, many methods have been proposed to detect faces. Roughly, they can be divided into two categories: feature-based and image-based. In the first category, the apparent properties of the face (such as facial features, edges, etc.) are exploited. Typically, the detection algorithms extract features from the image and then manipulate distances, angles and areas. Unlike the feature-based approaches, the image-based techniques apply a window-scanning algorithm on the image and classify the extracted sub-window into face and non-face classes. Neural networks, PCA, and SVM are examples of image-based techniques. For a new survey on face detection, see [1]. Majority of the proposed methods use image gray level values to detect faces in spite of the fact that most images today are color. As a consequence, most of these methods are computationally expensive and some of them can only deal with the frontal faces with little variation in size and orientation. To solve these problems, color-based face detection has recently become a new direction and exhibited a better performance.

Color-based face detection techniques first select skin-like regions and then detect or verify the presence of facial areas in these regions. Several approaches have been published and basically they differ in the modeling of the skin color (Terrilon *et al.* [5] recently presented a comparative study of these models under well behaved illumination conditions) and in the nature of feature extraction methods (most of the proposed approaches have considered the general problem of feature extraction where neural networks, edges, wavelets, templates, and histogram analysis are often used). For example, in [2], the face-like regions are detected using quantized skin-color region merging and wavelet packet analysis. This method, can detect faces with various sizes and different poses under minor illumination changes. However, it is also computationally expensive due to its complicated segmentation algorithm and time-consuming wavelet packet analysis.

In color-based face detection, the robustness of the skin color model is crucial to the overall system performance. The robustness of the model refers to its ability to detect skin color under varying illumination conditions. In previous work [3], we proposed a model for the skin color, which is robust under widely varying illumination conditions. In this paper, we extent this modeling to the face detection task. Thus, we present a face detection algorithm which uses the skin locus model, a robust method for extracting skin-like region candidates, and we perform the selection by simple but very useful techniques. We organized the filtering detectors in cascade to achieve high accuracy and keep our system simple and fast. While most color-based algorithms for face detection are based only on the extraction of facial features, our system can even deal with cases when these features are difficult to extract such as closed eyes. We demonstrate the ability of our method to detect faces under different conditions (size, orientation, and complex background).

2. Skin detection using skin locus model

Although different people have different skin color, but several studies have shown that the major difference lies largely in their intensity rather than their chrominance [5]. Several value distribution models have been compared in different color spaces (RGB, HSV, YCrCb, etc.)[1]. These distribution models have shown some efficiency in extracting skin-like regions under certain limited conditions. When only the chromaticity information is considered, also a relative robustness against intensity changes is achieved. However, this will not solve all the problems related to illumination and camera calibrations: skin chromaticities depend on the prevailing illumination and camera calibration light source. The more these two lighting factors differ, the bigger shift in chromaticities. Moreover, illumination color can be nonuniform over the face (in this case, even a proper calibration is not enough).

To solve these problems, we propose to use the skin locus which has performed well with images under widely varying conditions [3,4]. Skin locus (after Störring [6]) is the range of skin chromaticities under varying illumination/camera calibration conditions in NCC (normalized color coordinate) space as shown in Fig.1. In NCC space, intensity is defined as I=R+G+B and chromaticities are r=R/I, g=G/I and b=B/I. Because r+g+b=1, only the intensity and two chromaticity coordinates are enough for specifying any color uniquely. We considered r-b coordinates to obtain both robustness against intensity variance and good overlap of chromaticities of different skin colors. We used a quadratic function to define the upper bound of the r-b skin locus for a SONY camera (Fig.1.). The lower bound is defined by a five-degree polynomial function. Pixels with chromaticity (r, b) are labeled as skin or not using the constraint:

$$skin\ (r,b) = \begin{cases} 1\ if\ (\mathrm{b} > \mathrm{b}_d)\ \&\ (\mathrm{b} < \mathrm{b}_u)\ \&\ (\mathrm{r} \in [\mathrm{r1},\ \mathrm{r2}]) \\ 0 \quad otherwise \end{cases}$$

Where b_d is the lower bound, b_u the upper bound, r1=0.0877 and r2= 0.5645.

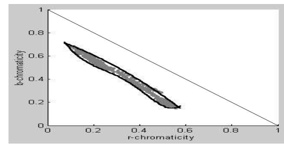

Figure 1. Appearance of the skin locus in r-b space for SONY camera (SONY dfw-x700)

3. Face detection algorithm

In order to detect faces, we firstly segment the face region candidates and then try to extract selected features (eyes, eyebrows and connected components) and verify their spatial relationships. This processing order allows a fast and robust analysis, because faces differ significantly from background by their color and shape. A different approach might start by detecting facial features and then infer the presence of a face.

We start by extracting skin-like regions using the skin locus shown in Fig.1. Using morphological operations (majority operator and applying dilations followed by erosions until the image no longer changes), we reduce the number of these regions. For every candidate, we verify whether it corresponds to a facial region or not. To increase the speed and robustness of our detector, we organized some operations on a cascade structure. The scheme shown in Fig.2 summarizes the filtering steps.

To deal with faces of different orientations, we firstly calculate the best ellipse fitting the face candidate. Based on the fact that the pixel value variations of other skin-like regions (such as hands) are smaller than those of face regions because of the presence of features with different brightness, we remove all face region candidates with pixel value variations smaller than a threshold. In order to improve the detection speed and achieve high robustness, we check the symmetry of the face and remove all the candidates when the symmetry is verified but no facial features are detected. Since it is not always possible to detect the facial features (due to different orientations, illuminations, etc.), we build a model of spatial arrangement of connected component features.

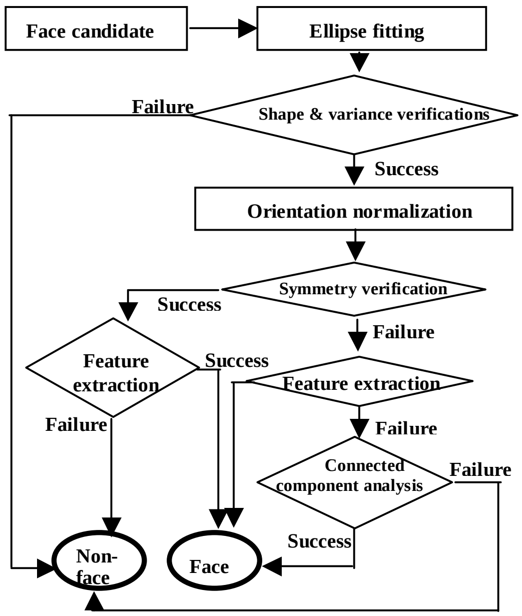

Figure 2. Detection scheme

3.1 Face candidate segmentation

The first step is to extract skin-like regions. We only keep the large connected components of skin color and remove isolated pixels corresponding to the background. However, not all detected skin regions contain faces. Some regions correspond to hands and others to exposed parts of the body, while some may correspond to objects with skin-like color (Fig.8).

3.2 Best-fit ellipse

We assume, which is obvious, that the overall shape of a face can be approximated by an ellipse.

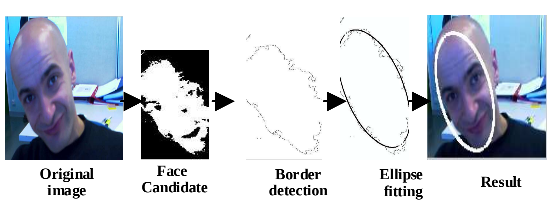

Figure 3. Example of ellipse fitting

To find the best-fit ellipse, we can either use an approach based on regions or on edges. In our case, we compute the best-fit ellipse based on the borders of the connected component corresponding to the face candidate. For this purpose, we use the least square ellipse fitting algorithm proposed by M. Pilu [7]. The algorithm considers only one ellipse that fits all the points. Thus, it is robust to noise and quite fast. Fig.3 shows an example of ellipse fitting.

3.3 Shape and variance verifications

The goal of this step is to reduce the number of false face candidates. After the best-fit ellipse operation, we only keep candidates in which the ratio of the major axis to the minor axis is within a certain range (in our experiments, we fixed the range to [1.11, 3.33]). Based on the observation that pixel value variations of facial regions are always more significant than those of other parts such as hands (which is due to the fact that features such as eyes, mouth and eyebrows have different color than skin), we remove all the candidates with a variance smaller than a threshold. Since facial features are better represented in the red channel, we consider only this channel in computing the variance. Due to illumination changes and other factors, not all hands will be removed and also we keep the threshold very small to not remove facial areas.

3.4 Orientation normalization

After computing the best-fit ellipse for each candidate, we perform image rotation to normalize the orientation of the face. Fig.4 shows an example of orientation normalization (the ellipse fitting step is shown in Fig.3).

If $(x_{rotated}, y_{rotated})$ denote the rotated coordinate of a pixel$_{(x,y)}$, then the coordinate transformation is defined by:

$$\begin{cases} x_{rotated} = x\cos\theta + y\sin\theta \\ Y_{rotated} = y\cos\theta - x\sin\theta \end{cases}$$

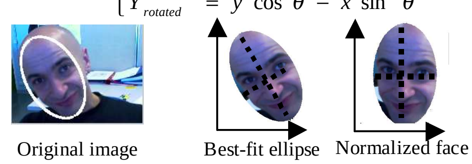

Figure 4. Orientation normalization.

3.5 Symmetry verification

We found in our experiments that when the symmetry of the face is verified, it is easier to detect the features. Thus, we implemented an idea of discarding all candidates when the symmetry is verified but no facial features have been found. After normalization, we compute a symmetry measure (SD) between the left and right parts of the face. We use 3x3 nonoverlapping windows to scan both parts. For every 3x3 window, we report local symmetry if the difference between the means of the pixel values corresponding to given 3x3 windows in both parts is smaller than 8. The SD corresponds to the ratio of number of local symmetries to the number of scanned windows. If more than 75 % of the 3x3 windows verify the local symmetry, then we consider that the face candidate is globally symmetric.

3.6 Eye and eyebrow detection

Given the face candidate after rotation, we consider the green channel in the interior of the connected component (experiments have shown that the green channel discriminates better the features we are looking for in the gradient images). Since the eyes and eyebrows are located in the upper half of the face, we only consider this region of the face. We calculate the gradient of the image on x dimension, then we determine the y-projection by computing the mean value of every row in the gradient image. Analyzing the y-projection, we found that the maximum corresponds to the horizontal position of the eyebrows (as shown in Fig.5).

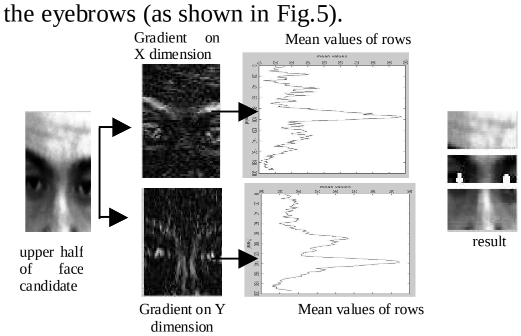

Figure 5. Detection of eyes and eyebrows

Having obtained the horizontal position, the vertical position is then determined by the x-projection. The x-projection is computed by averaging the pixel values around the 3 pixel neighborhoods of the horizontal position. In case of eyes, we proceed by the same way on the gradient image computed now on y dimension. Once we obtained the positions of the eyes and eyebrows, we verify their spatial relationships. The horizontal position of the eyes should be below the horizontal position of the eyebrows. Also, the ratio of the vertical eye-eyebrow distance to the face size should be within a certain range (we fixed it to [0.1,0.25]).

3.7 Connected component analysis

This step is only done in case of failure in the detection of eyes and eyebrows. Consider for example the image shown in Fig.6. The algorithm fails to detect the features. Since face contains eyes, mouth and eyebrow with darker color than skin, empty areas should exist inside the connected component of the face. We determine the best model (Fig.7) representing the connected component by computing five distances $D_{i(i=1..5)}$. We report a presence of face when the distance $D = \max_{i=1..5}(D_i) > threshold$ and also outside the feature areas, the number of non skin-like pixels is smaller than a defined threshold.

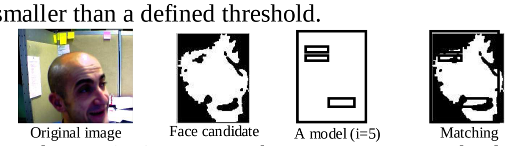

Figure.6. Connected component analysis

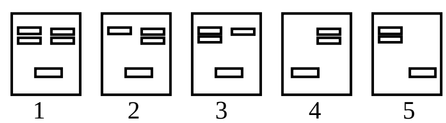

Figure.7. Different models of face

4. Experiments

To assess the robustness of our approach, we applied the system to detect faces in uncontrolled environment with complex background and natural illuminations.

Fig.8 shows an example with two face candidates (face and hand). After variance and symmetry verifications, both regions have been selected. Successful detection of eyes and eyebrows has been reported for the first candidate. After failure of feature detection in the hand region, the connected component analysis step rejected the false candidate.

Fig.9 shows some detection examples performed by the system under different conditions. Analyzing different filtering detectors and according to statistics obtained from a set of 500 detected faces, we found that in most successful detections (56%), the symmetry was not verified and then the features were detected. Only few face candidates (18%) verified the face symmetry. In 26% of the cases, the detection needed connected component analysis.

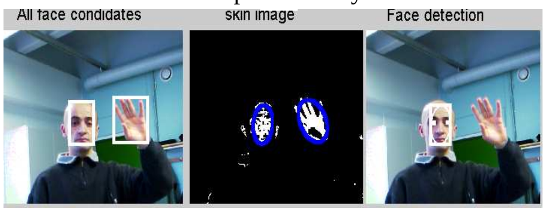

Figure.8. Face detection (with feature extraction)

Figure.9. Examples of successful detections

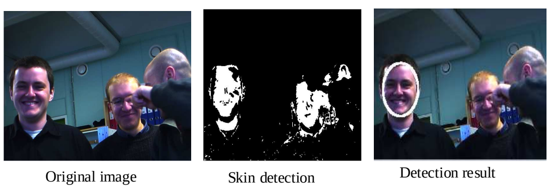

Figure.10. An example of failure

However, the system failed to detect faces when they are broken into small pieces of skin color due to occlusions. Fig.10 shows failure detection because the face skin is merged into a hand. The same failure is observed when two faces are too close.

Analyzing the importance of the different steps in removing the false face candidates, and according to statistics obtained from a set of 800 rejected candidates, we found that variance verification allowed 13% of removing and in 9% of the cases it is due to feature detection failure with successful symmetry verification. Considering these 800 cases, the shape verification has allowed 79 false candidate removing (about 10%). In 68% of the cases, the false face candidates passed through all the detectors.

5. Conclusion

A new architecture for face detection in color images under varying conditions was presented. First, skin-like regions are detected using skin locus model. Then, successive filtering detectors are applied to remove all regions which are not faces.

Although some false positive and false negative detections are found, the simplicity and robustness of the system are significant. The low complexity of the filtering detectors allows a real-time utilization of the approach. By using additional motion information as in [8], a further improvement in performance can be expected.

Under widely varying illuminations, the skin locus model has performed well in skin region extraction. However, this modeling is camera specific. To allow a more quantitative performance evaluation, a color face database for face detection should be collected, in which all the environment conditions and camera characteristics are included.

Acknowledgments

This research was sponsored by the Academy of Finland and the Finnish Graduate School in Electronics, Telecommunications and Automation (GETA).

References

[1] M. H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey", *in IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001, Vol. 24, No. 1, pp. 34-58.

[2] C. Garcia, and G. Tziritas, "Face detection using quantized skin color merged regions and wavelet packet analysis", *in IEEE Trans. Multimedia*, 1999, Vol. 1, No. 3, pp. 264-277.

[3] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Skin detection in video under changing illumination conditions", *In 15th Int. Conf. on Pattern Recognition*, Barcelona, Spain, September 3-8, 2000, Vol. 1, pp. 839-842.

[4] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietikäinen, "A physics-based face database for color research", *Journal of Electronic Imaging,* 2000, Vol. 9, No. 1, pp. 32-38.

[5] J. C. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human face in color images", *in Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2000, pp. 54-61.

[6] M. Störring, H. Andersen, and E. Granum, "Skin color detection under changing lighting conditions", *7th Symposium on Intelligent Robotics Systems*, Portugal, 20-23 July, 1999, pp. 187-195.

[7] M. Pilu, A. Fitzgibbon, and R. Fisher, "Ellipse-specific direct least-square fitting", *IEEE Conference on Image Processing*, Lausanne, 1996, vol. 3, pp. 599-602.

[8] H. P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multi-Modal System for Locating Heads and Faces",", *in Proc. 2nd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 1996, pp. 88-93.