FACE DETECTION FROM COMPLEX SCENES IN COLOR IMAGES

Doina Petrescu and Marc Gelgon
NOKIA Corporation Research Center
P.O.Box 100, FIN-33721, Tampere, FINLAND
Tel : +358 40 743 1959
e-mail : marc.gelgon@nokia.com

ABSTRACT

Automatic detection of faces in color images is a task of increasing importance for different multimedia applications. In this paper a 3-stage method for detecting human faces from images using color, shape and the gradient of luminance is presented. The aim of the method is to detect faces for a great variety of face sizes, positions and 3D orientations, including frontal and non-frontal views and situations when more than one face is present in the image. A measure of confidence that a region is a face is defined using parameters computed from the color, shape and local gradient of luminance features. This measure allows tuning the face detector to application requirements.

1. INTRODUCTION

The present paper presents an image analysis technique for detecting (and locating) human faces in color still images, whether there are none, one or several faces. The aim is to achieve this task for a wide and challenging range of cases (face positions, 3D orientations and sizes).

Different approaches have been used to solve the automatic face detection problem. The best known ones focus on gray-scale images. They can be based on eigenvectors decomposition of a face template and using neural networks or template matching to identify the best matches (location and scale) of face templates on the image [6]-[8]. These approaches are computationally expensive and sensitive to variations in illumination conditions, or head position and face orientation. Another group of approaches include shape-feature analysis where geometrical characteristics are used and candidate facial features are detected, spatial relationships are described by graphs and faces are found using graph analysis [9]. More recently, many authors have used color as an important cue to the presence of a human face [1]-[5]. Different color spaces were used to characterize skin color, and different type of color models were considered. However, experimental results show that even though the color feature can efficiently reduce the area of the image where face candidates may be found, it cannot be used alone for face detection with acceptable results. Another important cue about a face is the shape, which has been recently used in combination with color to detect faces [1]-[3].

In our case, it was needed to devise a method that can handle a wide variability of sizes at which a face appears, as well as far-from-frontal faces and some degree of occlusion. Fig.1 provides an overview diagram of the proposed scheme, composed of three successive stages. In the remainder of this paper, we present each of these three stages (sections 2,3 and 4), and provide experimental results (section 5).

2. COLOR-BASED REGION EXTRACTION

In the first stage, regions which are homogeneous in color and have high skin-color probability are separated from the image. First, a skin-color likelihood image is computed from the input image using a skin-color model. The hue-saturation (HS) color space is used for calculating the skin-color likelihood, being considered more appropriate to human perception of colors and in this sense it can provide superior results compared to other color spaces [2,3]. A Gaussian model is used for modeling the skin color probability. The model parameters were trained using skin samples collected from images obtained with different cameras, under various illuminations. Then, several iterations are performed consisting in skin-color likelihood image thresholding and non-linear processing of the result in order to find regions which are homogeneous in color and have high skin probability. The purpose of using several iterations is to remove the drawback of setting one single threshold on the skin likelihood may result in broken regions or in large skin-like blobs which contain different objects [1,4]. A Gaussian model is used for modeling the skin color probability. The model parameters were trained using skin samples collected from images obtained with different cameras, under various illuminations. Then, several iterations are performed consisting in skin-color likelihood image thresholding and non-linear processing of the result in order to find regions which are homogeneous in color and have high skin probability. The purpose of using several iterations is to remove the drawback of setting one single threshold on the skin likelihood may result in broken regions or in large skin-like blobs which contain different objects [1,4]. In this procedure the thresholds are set using the histogram of the skin likelihood image, in decreasing order (the highest threshold first, to select only the most likely skin pixels). The result of thresholding is filtered with a nonlinear filter (the 5x5 was used in the experiments) to remove small groups of pixels and fill in small holes. Regions which represent ‘nuclei’ of homogeneous colored areas are obtained.
The borders of the uniform colored regions need to be found in the color domain, using local color information, which was disregarded by thresholding. A dilation procedure, denoted by *dilation constrained by color homogeneity of the region* is defined and customized to correct the result of thresholding and find color region borders. Pixels from the border are added to the region as long as they fulfill a color constraint. The constraint used for dilation is the absolute color difference between the average color in the neighborhood of the pixel to be added and the average color of the region. At the end of each iteration, pixels which belong to detected regions are removed from the skin likelihood image, to simplify the region detection for the next (lower) threshold. In this way regions detected in different iterations do not overlap.

After each iteration a new threshold is calculated using the histogram of the remaining pixels (pixels which do not belong to detected regions). This stage may lead to oversegmentation of uniform colored regions. The second stage aims at region-level analysis to alleviate this issue.

### 3. MERGING OF EXTRACTED REGIONS

In the second stage, adjacent elementary regions are merged based on a criterion which jointly combines color similarity and an elliptical shape constraint. The color similarity measure between two regions, \( R \) and \( P \), denoted by \( C_{\text{color}}(R,P) \), is calculated using the absolute difference between their average colors. The shape criterion, denoted by \( C_{\text{shape}} \) is a measure on how well the shape fits an ellipse. This measure, proposed in [3], is the ratio between the number of pixels erroneously represented by the ellipse (the 'holes' inside the ellipse and pixels from the region which are outside the ellipse) and the total number of pixels of the ellipse. The combined color and shape criterion is defined as (1).

\[
C_{\text{merge}}(R,P) = C_{\text{color}}(R,P) \cdot C_{\text{shape}}(R \cup P) \quad (1)
\]

This criterion has small values when regions have similar colors and the shape of the region obtained by merging fits well to an ellipse. For each region \( R \), the joint color-shape criterion \( C_{\text{merge}} \) with every neighboring region \( P \) is computed. The region is merged with the neighbor for which the lowest value of the joint criterion is obtained. Regions are merged only if the shape resulted by merging fits better to an ellipse than each of the merging parts (2).

\[
C_{\text{shape}}(R \cup P) < \min (C_{\text{shape}}(R), C_{\text{shape}}(P)) \quad (2)
\]

Pairwise merging continues for all regions until no more merging can be done. Regions obtained after the first two stages are included in a candidates face list.

### 4. SELECTION OF OBTAINED REGIONS

In this last stage, parameters computed from color, shape, and the gradient of luminance inside the region, are used to evaluate the candidates. Regions which are too small to reliably represent faces (less than 2% of image size in the experiments) are first removed from the candidates list. Besides, regions whose shape is too far from an ellipse, or whose ratio between the major and the minor axis does not reasonably fit to facial aspect ratio, having allowed for a variety of viewpoints, are removed also from the candidates list. In deciding whether a region is a face, a simple texture statistics is employed, namely the percentage of pixels in the region having the gradient of luminance along the major axis of the region higher than a threshold predetermined from experiments. It has shown an efficient discriminant against many false alarms. A combined criterion denoted by *face cost factor* is defined and calculated for each region from the average color, shape criterion and the gradient-based statistics.
described above, so that poor but acceptable values for one of the criteria can be made for by other criteria. The face cost factor has low values when parameters computed from features show a high probability that the candidate is a face. The region having the lowest cost factor is denoted by reference face of the image. A confidence level with values in (0,1] is defined using the face cost factor to show the probability of a candidate being a face with respect to that of the reference face. By deciding to accept only candidate faces whose confidence level are above a certain threshold, the user may tune the face detector to the application. He may notably choose between the selection of only those candidates which have very high probability of being faces, and the acceptance of many candidates from the list as face regions.

5. EXPERIMENTAL RESULTS

Color-learning and testing sets of images were built, that represent a large variety of cases. Experimental results are provided in Table 1.

216 images containing one face in different positions and under various illumination conditions. The confidence level was set to 1, such that only the reference face is detected in each image. In 162 images faces were correctly detected. In 5 images no face was detected and in 49 images objects or regions of the background were erroneously shown as faces. Most of the problems were caused by failures in merging the different parts of the faces together, in the second stage, because the simultaneous merging of more than two regions is not considered.

31 images containing 87 faces were used to evaluate the face detection method in images containing more than one face. The tests were performed with 3 different confidence level thresholds: 0.2, 0.5 and 0.9. The experiments illustrate how the confidence level threshold allows the user to tune the face detector to the application requirements. Setting a low threshold (close to 0) for the confidence level leads to a very high detection rate, but many false alarms. This is useful when the number of faces is not known, but it is desirable that faces are not missed. Setting a high threshold for the confidence level (close to 1) usually results in one face at most detected per image. If objects from the background have skin-like color and elliptical shape, it may happen that other regions than faces are detected. However, setting a high threshold has shown a good choice for neighbouring issues: if the user only needs to detect whether at least a face are present, or the user knows that there is one and only one face, which has to be localized.

Three images representing the variety of situations encountered in the evaluation and the face detection results when the confidence level threshold is 0.2 are presented in Fig.2. Computational time remains low (1 to 2 seconds on a standard PC).

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Images</th>
<th>Faces</th>
<th>Faces detected</th>
<th>False detections</th>
<th>Missed faces</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Images containing one face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>216</td>
<td>216</td>
<td>211</td>
<td>49</td>
<td>5</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Images containing more than one face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>31</td>
<td>87</td>
<td>98</td>
<td>40</td>
<td>11</td>
<td>58</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td>62</td>
<td>16</td>
<td>41</td>
<td>46</td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td>31</td>
<td>4</td>
<td>60</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 1. Face detection results
6. REFERENCES


Fig. 2: original images (up) and faces detected (down) when the confidence level threshold is 0.2