Face Detection on Still Images Using HIT Maps[*]

Ginés García Mateos[1], Cristina Vicente Chicote[2]

[1] Dept. Informática y Sistemas,
University of Murcia, 30.170 Espinardo, Murcia, Spain
ginesgm@um.es
[2] Dept. Tecnologías de la Información y las Comunicaciones
University of Cartagena, 30.202 Cartagena, Murcia, Spain
cristina.vicente@upct.es

Abstract. We present a fully automatic solution to human face detection on still color images and to the closely related problems of face segmentation and location. Our method is based on the use of color and texture for searching skin-like regions in the images. This is accomplished with connected component analysis in adaptatively thresholded images. Multiple candidate regions appear, so determining whether each one corresponds or not to a face, solves the detection problem and allows a straightforward segmentation. Then, the main facial features are located using accumulative projections. We present some results on a database of typical TV and videoconference images. Finally, we extract some conclusions and advance our future work.

1 Introduction

Most of the existing techniques for face detection suffer from being either quite expensive or not very robust. In the first group, we can find systems that are based on exhaustive multiscale searching using neural networks [4] or eigen-decomposition [3] and, usually, color is not used. On the other hand, systems that make use of color features [5], [6], are computationally less expensive but are not very robust and present serious problems under uncontrolled environments.

The research described in this paper deals with the problem of human face detection on color images and the closely related problems of face segmentation and facial features location. We propose a technique based on color features which is intended to work under realistic uncontrolled situations. It has been tested using a database of images acquired from TV and from a webcam, achieving very promising results.

The key point in face analysis using color images is to search and describe skin-like regions [6]. We have defined a representation space named HIT (Hue, Intensity and Texture), that allows a simpler detection of skin-like regions. A fast connected component labeling algorithm is applied on thresholded HIT images, using adaptative thresholding in order to achieve invariance to clutter, noise and intensity.

[*] This work has been partially supported by Spanish CICYT project TIC1998-0559.

2 Skin-Region Searching in HIT Maps

2.1 HIT Maps

Differences among typical images of human skin (due to environment conditions, the acquiring system or the skin itself) do mainly cause intensity variations [7]. Thus, the color spaces most widely used for skin analysis are designed to be intensity invariant. We can mention the chromatic color space, or normalized *(r, g)* [7] and the HSV or HSI spaces [2],[5],[6], among others. But in many non-trivial cases, color features are not enough to separate skin from background objects. In these cases, intensity gradient is useful to detach face from other objects, and intensity itself may also be useful.

We have defined a representation space, named HIT, so that each RGB input image is transformed, in a preprocessing stage, into a three-channel image: *Hue*, *Intensity* and *Texture*. This *Texture* is defined as the magnitude of the multispectral gradient in RGB using the Sobel operator. The use of this particular color space transformation is justified in detail in [2]. Fig. 1 shows two sample images used in the tests, and their corresponding HIT transformations.

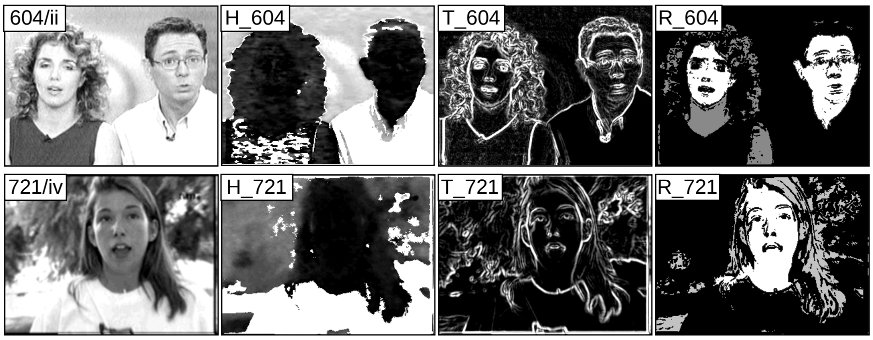

Fig. 1. Skin-region searching in HIT maps. From left to right: input image; hue channel; texture channel; skin regions found using adequate thresholds (those verifying size and shape criteria, in white).

2.2 Skin-Region Searching

A classification process is defined on HIT patterns, so that each vector *v = (h, i, t)* is classified into one of two classes: skin or non-skin. We use a simple thresholding on the three channels, thus simplifying the training and classification. Contiguous pixels that are classified as skin patterns are then joined into skin-like regions, using a connected component labeling algorithm. This algorithm can be implemented with a single and very efficient scan of the image, from left to right and from top to bottom.

This method works quite well when adequate thresholds are selected, as in Fig. 1. But these thresholds may change from one image to another, so they can not be a priori fixed. Fig. 2 shows a sample of the variety that skin color and texture may undergo in different images, due to the environmental lighting conditions, the acquiring system and the noise in the video signal.

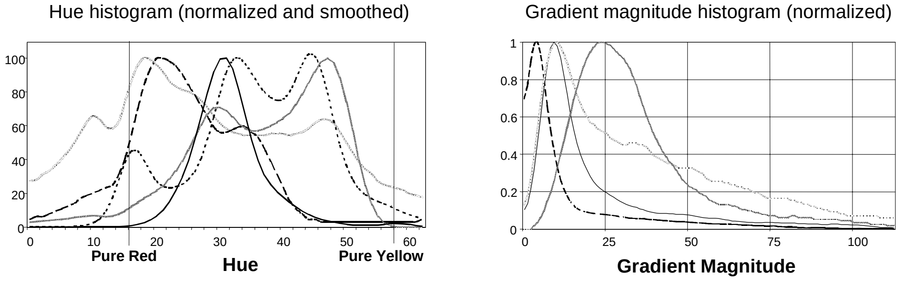

Fig. 2. *Hue* and *Texture* channel histograms for some skin-color regions in various images.

Our proposal is to use these histograms in order to adapt the color and texture models for each image in particular. For the color model, the hue of the skin corresponds to a maximum in that histogram, lying between pure red and pure yellow. For the texture and intensity channels, thresholds are calculated using the corresponding histograms. These thresholds can be more or less restrictive, so different adaptative tests appear. The whole process of skin-region searching is depicted in Fig. 3.

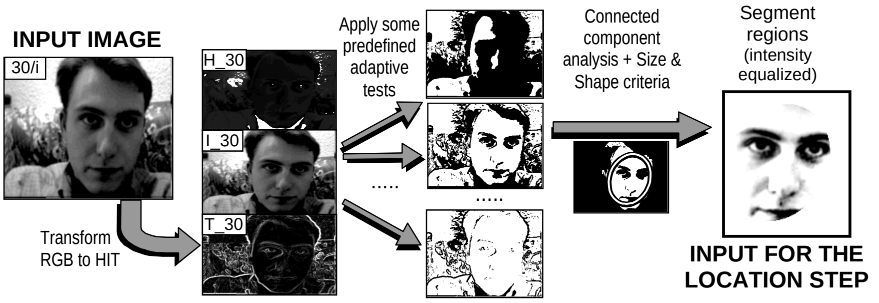
INPUT IMAGE
30/i
Transform RGB to HIT
H_30
I_30
T_30
Apply some predefined adaptive tests
.....
.....
Connected component analysis + Size & Shape criteria
Segment regions (intensity equalized)
INPUT FOR THE LOCATION STEP

Fig. 3. Candidate skin-region searching process. The output are segmented candidate regions.

First the input image is transformed into an HIT map. Using the histograms, some adaptative tests are defined, that result in binarized images. In the experiments, six of these tests have been applied to each image. Connected components are then searched in binarized images. We apply shape criteria on the resulting regions to select those ones with elliptical shape that are not very elongated. These criteria are described in more detail in [2]. Finally, the regions that satisfy both criteria are segmented, equalizing the intensity channel with a linear model of the skin-region intensity, as in [4]. The segmented area corresponds to the ellipse that better fits the candidate region.

3 Facial Features Location

From the segmented regions obtained through the previous steps, the main facial features (i.e. eyebrows, eyes, nose and mouth) can be accurately located by analysing its horizontal and vertical integral projections. Besides, the a priori knowledge about the human facial structure, makes it possible to apply some heuristics in order to guide the location process in an efficient and smart way.

Integral projections on edge images and on intensity images [1], [6], have proved to be useful for facial features location. Given a segmented grayscale input image *I(x,y)*, its horizontal and vertical integral projections are defined as $HP(y) = \Sigma I(\cdot,y)$ and $VP(x) = \Sigma I(x,\cdot)$. These projections are smoothed in order to remove some small spurious peaks. Gaussian masks of different size are used for this purpose, depending on the size and contrast of the input image. Then, the location of the facial features can be obtained from the local maxima and minima extracted from these softened projections, as shown in Fig. 4. If no prominent peaks are found in the expected positions, then we infer that there is no face in the image. This detection test implies that a face exits when all its features are located. This way, the number of false-positive errors (see Table 1) is very small. But many existing faces are difficult to be located, so we can also consider that a face is detected when a candidate region is found. We will denote these two possibilities as detection *before location* and *after location*.

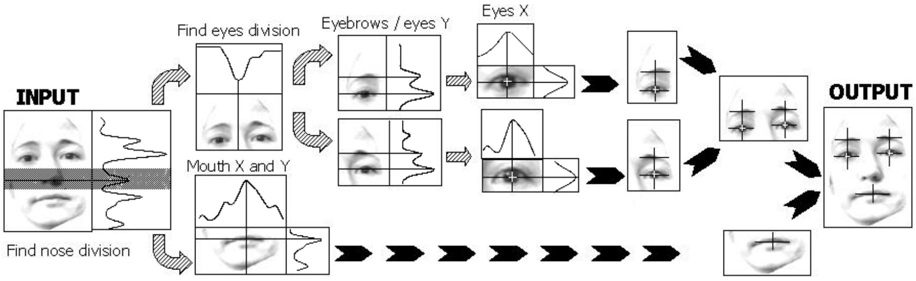

Fig. 4. Facial components location. The segmented region is successively divided into smaller ones by finding the maxima and/or minima within their softened integral projections and by applying some a priori knowledge about the face geometry.

4 Experimental Results

For our experiments, we have constructed a database of color images containing human faces within complex environments. Most of the existing face databases used for face detection are composed by gray scale images [4]. On the other hand, color face databases, used for person recognition, are not adequate for detection benchmarks. At the moment, our database is composed by 195 color images, some of them not corresponding to faces or containing more than one. A total of 101 distinct individuals appear on them, out of 199 existing faces. All these images have been acquired from a webcam and from 27 different TV channels, a few of them rather noisy.

The images have been classified into six groups: i) videoconference images; ii) news with presenters' faces and shoulders only; iii) news with presenters' faces and busts; iv) TV series, documentaries and outsides; v) fashion shows; and vi) non-faces. The detection and location results achieved by our system, are shown below.

Table 1. Face detection and location results

Image group	Existing faces (images)	Faces detected before / after location	False-positive before / after location	False-negative before / after location	Location accuracy error
i)	35 (35)	34 / 30 97.1% / 85.7%	0 / 0 0% / 0%	1 / 5 2.9% / 14.3%	4.1 %
ii)	14 (14)	13 / 9 92.9% / 64.3%	1 / 0 7.1% / 0%	1 / 5 7.1% / 35.7%	3.6 %
iii)	58 (55)	53 / 35 91.4% / 60.3%	21 / 0 38.2% / 0%	5 / 23 8.6% / 39.7%	2.4 %
iv)	78 (69)	63 / 30 80.8% / 38.5%	27 / 4 39.1% / 5.8%	15 / 48 19.2% / 61.5%	1.9 %
v)	14 (13)	11 / 10 78.6% / 71.4%	3 / 0 23.1% / 0%	3 / 4 21.4% / 28.6%	6.2 %
vi)	0 (9)	0 / 0 - / -	2 / 0 22.2% / 0%	0 / 0 - / -	-
TOTAL	199 (195)	174 / 114 87.4% / 57.3%	54 / 4 27.7% / 2.1%	25 / 85 12.6% / 42.7%	3.1 %

A comparison of some state-of-the-art methods related to face detection can be found in [4], where the detection rates exhibited are between 78.9% and 90.5%. Compared to them, our method achieves similar results but requires less expensive computations. The performance highly varies from the most favorable group, that of videoconference images, with 85.7% of the faces located, to the more difficult one, group iv), with only 38.4% of the faces located. However, the false-positive rate *after location* is always very low, with a total of 2.1%.

The location results have been compared with a manual measure of the facial features positions. The location accuracy with respect to the face height, is nearly always below 4%, with a mean accuracy error of 3.1%. Some results of our detection method are shown in Fig. 5. Note that a faced is said to be detected *before location* when a box appear, and *after location* when all the facial features (marked with +) are found.

5 Conclusions and Future Work

The method here described, offers a unified solution to three basic problems of face analysis: detection, segmentation and location. We have defined the HIT representation space, which takes into account color, intensity and texture information. Adaptative techniques are used for establishing color and texture models of the images. Then, the main facial features are located by searching certain local maxima in the integral projections of intensity images, given some a priori geometrical constrains.

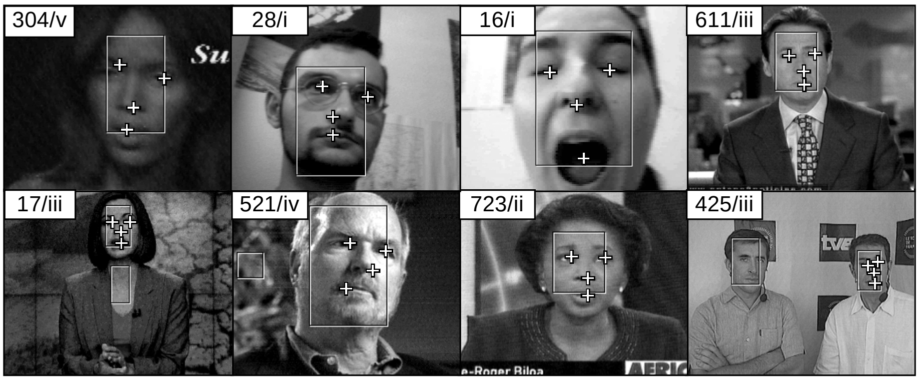

Fig. 5. Face detection results in some sample images. Candidate regions are indicated with a bounding box. Eyes, nose and mouth are marked if they have been located.

The achieved results are very promising, with a total detection ratio *before location* of 87.4%. This performance decreases when the location step is considered. This is due to a bad segmentation, which includes hair or neck as a part of the face region. We are currently working on improving segmentation, but it is worth to note the very low false-positive error obtained (2.1%) with respect to the number of images.

Our future work also includes the application of our method to person identification, facial expression recognition, videoconference coding and the extension of our approach to face tracking in image sequences.

References

1. Brunelli, R., Poggio, T.: Face Recognition: Features versus Templates. IEEE Transactions on PRIA, Vol. 15, No. 10 (October 1993) 1042-1052
2. García, G., Vicente, C.: A Unified Approach to Face Detection, Segmentation and Location Using HIT Maps. SNRFAI'2001, Castellón de la Plana, Spain, (May 2001)
3. Moghaddam, B., Pentland, A.: Probabilistic Visual Learning for Object Detection. International Conference on Computer Vision, Cambridge, MA (1995)
4. Rowley, H.A., Baluja, S., Kanade, T.: Neural Network-Based Face Detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1 (January 1998) 23-38
5. Sigal, L. et al.: Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. IEEE Conference on CVPR (2000)
6. Sobottka, K., Pitas, I.: Looking for Faces and Facial Features in Color Images. PRIA: Advances in Mathematical Theory and Applications, Vol. 7, No. 1 (1997)
7. Yang, J., Waibel, A.: A Real-Time Face Tracker. In Proceedings of the Third IEEE Workshop on Applications of Computer Vision (1996) 142-147