A Unified Approach to Face Detection, Segmentation and Location Using HIT Maps

Ginés García Mateos†, Cristina Vicente Chicote‡

†Dept. Informática y Sistemas
University of Murcia, 30.070 Espinardo, Murcia, Spain
ginesgm@um.es

‡Dept. Tecnologías de la Información y las Comunicaciones
University of Cartagena, 30.202 Cartagena, Murcia, Spain
cristina.vicente@upct.es

Abstract

In this paper, we present a unified approach to the problems of human face detection, segmentation and location. Color and texture are used for searching skin-like regions in the images. Determining whether each region corresponds or not to a face solves the detection problem, and allows a straightforward segmentation. Then, region information is also used for locating the main facial features. Using this unified approach increases the efficiency of solving the three problems, and allows us to focus on the skin region searching, so any improvement will affect the global performance. We present some results and conclusions and advance our future work.

Keywords: color and texture segmentation, face analysis, detection and location.

1 Introduction

Nowadays, the research in perceptual interfaces and intelligent environments is becoming of increasing importance in the computer vision community [1]. A new generation of applications has arisen that is based on intuitive, simple and natural human interfaces, referred by some authors as the "looking-at-people" field. Although most techniques are application dependent, a number of common problems appear in the "looking-at-people" field. Focusing on human face analysis on still images, we can mention [4-7]: face detection (finding the number of faces that appear in an image), face segmentation (determining the area in the images occupied by each face) and face location (accurately positioning the face and its main features). In this paper, we propose a unified approach to solve the three problems stated above, which allows us to deal with them in a common framework of region searching, in a very efficient way.

2 The Unified Framework: Region Searching in HIT Maps

We believe that the key point to unify face analysis (i.e. detection, segmentation and location), is to search and describe skin-like regions. For this purpose, we define a representation space named HIT (Hue, Intensity and Texture), that allows a simpler detection of skin regions, using a connected component labeling algorithm.

2.1 HIT Maps

Color features have already been used for face, hand or skin detection, location and tracking [2, 3, 5, 7]. Human skin color is clustered in certain regions of the color space, as already shown by Yang and Waibel in their work about human skin color [2]. Differences between individuals are mainly due to distinct melanin concentration, which causes intensity variations. Two color spaces are the most widely used for skin analysis: the chromatic color space, or normalized *(r, g)* [2, 7] and the HSV space [3, 5].

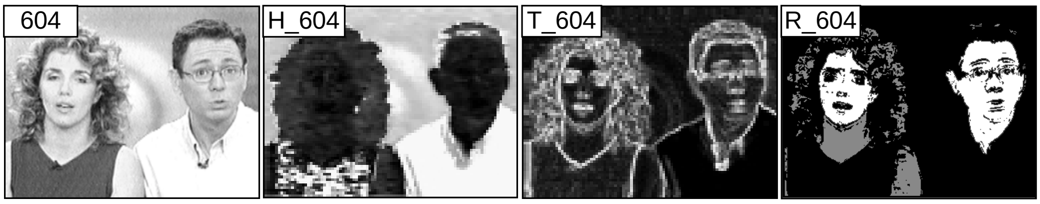

Figure 1: Region searching in HIT maps: input image; hue channel; texture channel; and skin regions found (those verifying size and shape criteria, in white).

Our proposal is to use only the *Hue* component for color analysis. Using only this channel simplifies the training and classification of color patterns, while retaining the substantial chromatic information of RGB. Besides, its computation can be optimized by using small size precalcutated tables [3]. On the other hand, adding the *Intensity* channel, it is possible to define a test for rejecting pixels that clearly do not belong to skin regions. The use of this channel compensates the instability of the *Hue* for dark and light colors.

However, color information is not enough for discriminating skin regions, as any arbitrary background object may have a skin-like color. Human skin has a uniform texture, so different skin-color regions can be separated using gradient information. Thus, *Texture* is the third channel of our representation space. It is defined as the magnitude of the multispectral gradient in RGB. This gradient magnitude is calculated using the Sobel operator on the three channels, and the resulting values are combined with a maximum.

2.2 Region Searching

Input RGB images are transformed into HIT maps. Then, a classification method is defined on HIT patterns, so that each vector $v = (h, i, t)$ is classified into one of two classes: skin or non-skin. Contiguous pixels that are classified as skin patterns are then joined into skin-like regions, using a connected component labeling algorithm. This algorithm is implemented with a single and efficient scan of the image, top to bottom and left to right.

During the component labeling process, a description is updated for each region found R, which contains: mean M_R, covariance Σ_R, number of pixels N_R, and mean intensity gradient ∇I_R. This description may also include some other values incrementally calculated. The result of this process, i.e. number of regions R_{max}, their descriptions and the pixel/region associations $R(x,y)$, constitutes the common framework over which the problems of face analysis are solved. Figure 1 shows a sample input image acquired from TV, its HIT transformation and the regions found by the labeling algorithm.

3 Solving Face Analysis Problems

3.1 Face Detection

Let's consider the data constituting the framework of our approach: number and description of skin-like regions, and the pixels corresponding to each one. Face detection becomes a test applied on all the detected regions, that determines whether each one is a face or not. This test uses a priori knowledge about face geometry and expected size. For example, the following test based on region shape and size has shown good results:

if N_R>MIN_PIXELS **and** $\sigma_{YR}/\sigma_{XR} \in (ELO_{min}, ELO_{max})$ **and** $4\pi Sqrt(Det(\Sigma_R))/N_R \in (COM_{min}, COM_{max})$
then R IS A FACE

The ratio σ_{YR}/σ_{XR} indicates elliptical elongation of region R. Typical values for faces are between 1.22 and 1.90. The value $4\pi Sqrt(Det(\Sigma_R))/N_R$ measures the compactness of R, and it is in the range (1.16 - 1.60) for faces correctly detected. An example of the application of this test is shown in Figure 1. In our experiments, this test has shown a 12% false-negative error, and a 6% false-positive rate. However, we have to note that this is a preliminary detection step. The detection performance is highly improved with the results of face location, as we will discuss in subsection 3.3.

3.2 Face Segmentation

Face segmentation consists of determining which portion of the image belongs to each face. This segmentation is applied to those regions that verify the detection test. Here, we propose two methods for face segmentation: elliptical and polygonal. In the first one, the segmented area is an ellipse corresponding to the gaussian parameters of that region (M_R, Σ_R). This method is simpler, but generates a coarse approximation of the facial shape.

In the polygonal method, a list of points is first extracted by contour-following of R, using $R(x, y)$. Then, we use the IPE algorithm [8] to reduce the number of points of the polygon, minimizing the area suppressed by the removed vertices. Finally a convex criterion is used to remove vertices that generate convexities. Some results on face segmentation, using the polygonal method, are shown in Figure 3.

3.3 Facial Features Location

Accurately locating the main facial features (i.e. eyebrows, eyes, nose and mouth) can be achieved effortlessly within our framework, using the segmented images of the faces, their gaussian parameters (M_R, Σ_R) and the intensity gradient information ∇I_R. The approach we propose in this paper, depicted in Figure 2, is based on the analysis of the horizontal and vertical integral projections of the faces, segmented using the polygonal method. These projections are defined on intensity segmented images, equalized using ∇I_R to compensate for illumination changes on the faces, as shown in [6].

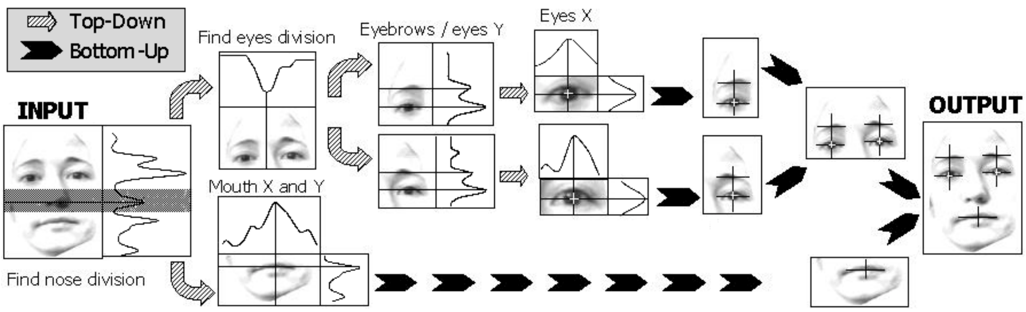

Figure 2: *Top-Down:* successively divide the segmented region into smaller ones using integral projections. *Bottom-up:* combine the locally computed data.

Integral projections on edge images [4], and on intensity images [5], have proven 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.

Then, the location of the facial features can be obtained from the local maxima and minima extracted from these softened projections, as shown in Figure 2. If no prominent peaks are found in the expected positions, then we infer that there is no face in the image. This is a refinement of the detection test given in 3.1, that highly increases its reliability.

Unfortunately, any rotation affects the performance of the integral projections method due to facial features overlapping. To partially undo this rotation in the original images, we use the gaussian covariance matrix Σ_R provided by the framework, which can be followed by an iterative horizontal alignment of the eyes when detected.

3.4 Face Analysis Results

The whole system has been tested on a set of 121 color images, acquired from TV and from a videoconference camera. Some of the images do not correspond to faces or contain more than one. Some examples and results are shown in Figure 3. Face detection, improved with the restrictions given in 3.3, achieves a false-positive error of 2%, and false-negative of 9%. The segmentation step works quite well in nearly all cases (see Figure 3). Only in the 5% of the samples some non-face areas are added to the segmented region. Facial component location is accurately computed if a proper segmentation is provided. Eyes, eyebrows and mouth are correctly located in 94% of the detected faces.

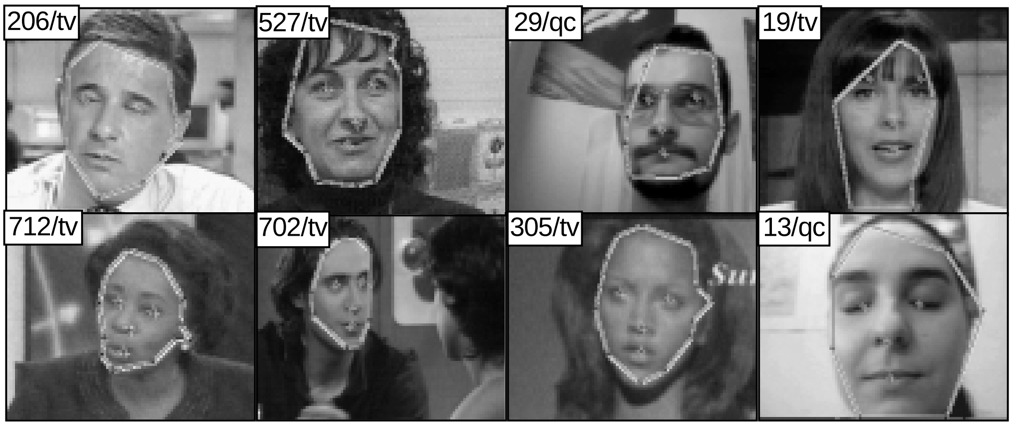

Figure 3: Examples of face analysis results: detection, segmentation and location.

A comparison of some state-of-the-art methods related to face detection can be found in [6], where the detection rates exhibited are between 78.9% and 90.5%. Some of these systems are based on neural networks and exhaustive searching, being quite inefficient. Other techniques [5], suffer from being not very robust. Compared to them, our method achieves similar results but in a simpler, more efficient and robust way.

4 Conclusions and Future Work

The method here described, offers a unified solution to three basic problems of face analysis: detection, segmentation and location. The key point is to find and describe skin-like regions in the images. This is achieved using the HIT representation space that takes into account color, intensity and texture information. Within the designed framework, face detection is accomplished by defining an adequate shape criterion. Face segmentation consists of convex region extraction for the areas that satisfy the detection criterion. Finally, the main facial features are located by searching certain local maxima in the integral projections of intensity images, given some a priori geometrical constrains.

This approach makes it possible to improve the solution of the three problems when any of them, particularly the skin region searching, is enhanced. Besides, it increases the efficiency of solving them independently and can be easily extended to solve other similar problems related to face analysis. Although this research is in a preliminary stage, the results obtained are very promising. Our future work includes the application of our method to person identification, facial expression recognition and videoconference coding. We are also working in the extension of our approach to face tracking in sequences.

References

[1] A. Pentland. Looking at People: Sensing for Ubiquitous and Weareable Computing. *IEEE Transactions on PAMI*, Vol. 22, No. 1, January 2000.

[2] J. Yang and A. Waibel. A Real-Time Face Tracker. In *Proceedings of the Third IEEE Workshop on Applications of Computer Vision*, pp. 142-147, 1996.

[3] L. Sigal et al. Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. *IEEE Conference on CVPR*, 2000.

[4] R. Brunelli, T. Poggio. Face Recognition: Features versus Templates. *IEEE Transactions on PRIA*, Vol. 15, No. 10, pp. 1042-1052, October 1993.

[5] K. Sobottka, I. Pitas. Looking for Faces and Facial Features in Color Images. *PRIA: Advances in Mathematical Theory and Applications*, Vol. 7, No. 1, 1997.

[6] H.A. Rowley, S. Baluja, T. Kanade. Human Face Detection in Visual Scenes. *TR CMU-CS-95-158R*. November 1995.

[7] R.J. Huang. Detection Strategies for Face Recognition Using Learning and Evolution. *Ph.D. George Mason University*, May 1998.

[8] A. Pikaz and A. Averbuch. On Automatic Threshold Selection for Polygonal Approx. of Digital Curves. *Pattern Recognition*, Vol. 29, No. 11, pp. 1835-1845, 1996.