Finding Faces in Color Images through Primitive Shape Features

Murad Al Haj*, Ariel Amato*, Xavier Roca*, and Jordi Gonzàlez+

* *Computer Vision Center, Department d'Informàtica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*
E-mail:malhaj@cvc.uab.es

+ *Institut de Robòtica i Informàtica Industrial (UPC-CSIC), Edifici U, Parc Tecnològic de Barcelona, 08028, Barcelona, Spain*
E-mail:jordi.gonzalez@cvc.uab.cat

Abstract Face detection is a primary step in many applications such as face recognition, video surveillance, human computer interface, and expression recognition. Many existing detection techniques suffer under scale variation, pose variation (frontal vs. profile), illumination changes, and complex backgrounds. In this paper, we present a robust and efficient method for face detection in color images. Skin color segmentation and edge detection are employed to separate all non-face regions from the candidate faces. Primitive shape features are then used to decide which of the candidate regions actually correspond to a face. The advantage of this method is its ability to achieve a high detection rate under varying conditions (pose, scale,...) with low computational cost.

Keywords: Face Detection, Facial Features, and Skin Color

1 Introduction

Detecting faces in images is essential in many applications of computer vision. Of these applications, facial expression analysis and human computer interaction have attracted much interest lately. The main challenges facing any face detection system include: pose variation, presence or absence of structural components (beards, mustaches, and glasses), facial expressions, occlusion, image orientation, and imaging conditions. Many techniques exist for solving these problems and trying to achieve a robust detection. A survey for face detection is presented, [1], where the different methods are classified into four, sometimes overlapping, categories: knowledge-based methods, feature invariant approaches, template matching methods, and appearance-based methods. Viola and Jones, [2], have introduced a rapid and robust appearance-based object detection method, using Haar-like features and Adaboost and applied it to face detection. This method has been gaining popularity among researchers. However, the problem with this method is that it is very sensitive to the face pose. In other words, it is very difficult to detect frontal faces and profile faces with the same cascade of weak classifiers. Therefore, two cascades, one for frontal and another for profile, need to be trained for proper classification; and even with these two cascades, some faces whose pose is between frontal and profile, 45 degree for example, can't be detected. Some modifications have been proposed to make this approach more robust for multi-view [3]. However, some researchers have returned lately to the use of color as a robust cue for face detection, [4] and [5], where the proposed methods are based on segmenting the image using human skin color. A method that detects face using the spatial arrangement of skin patches is available in [6].

In this paper, a novel face detection method based on skin color segmentation is presented. The main contributions of this approach can be summarized in proposing an efficient skin color model for image segmentation and using simple primitive shape features to achieve robust detection under varying poses and complex backgrounds.

This paper is organized as follows. An overview of the system is presented in section 2. Section 3 introduces our proposed skin color model while section 4 presents the face localization method. In section 5, experimental results are shown. Concluding remarks are discussed in section 6.

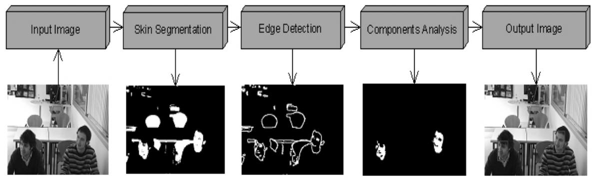

Figure 1: An overview of the system

2 System Overview

The complete system is shown in Fig. 1. It starts by segmenting the image into regions that contain possible face candidates, while those that do not contain a face object are dropped. This segmentation helps accelerate the detection process. Later, an edge detection filter is applied (i) to separate the candidate faces from any background component that has a color similar to the skin, and (ii) to disconnect foreground components, such as two touching faces. Next, the components are analyazed and some primitive shape features of the human face are used to decide which region is a face and which is not.

3 Skin Color Model

Experience proved that skin is an effective and robust cue for face detection. Color is highly invariant to geometric variations of the face and it allows fast processing. Different colorspaces were studied by different researchers in order to find an optimal representation for the skin color distribution. A survey about pixel-based skin color detection can be found in [7]. Many researchers claim that other colorspaces, such as YCbCr, [5], performs better than the RGB colorspace. However the authors of [8] argue that the separability between skin and non-skin classes is highest in the RGB colorspace and that dropping the illumination component worsens this separability. Many papers on skin detection do not properly justify the colorspace of their choice probably because it is possible to obtain good results, for a limited dataset, in any colorspace [7]. The performance of a skin detection method depends on a combination of the colorspace and the classifier, and we believe that it is possible to find an acceptable detector for any colorspace.

In this section, we present a classifier for skin color based in the RGB colorspace. The proposed

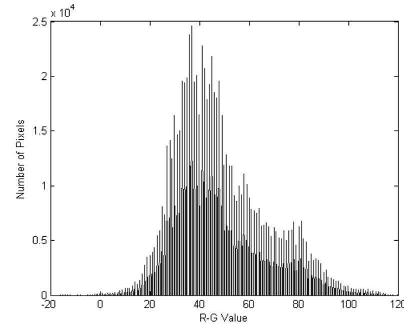

Figure 2: Distribution of R-G values in skin pixels

algorithm is shown below:

(R,G,B) is classified as skin if:
$20<R-G<90$
$R>75$
$R/G<2.5$

The advantages of this method are its simplicity and computational efficiency since no transformation is needed to go to another colorspace. It is based on an idea presented in [9], where it was noticed that when subtracting the red channel from the green channel, the skin pixels tend to have distinctive values-greater than the non-skin pixels. In our experiments, around 800,000 skin pixels where taken from 64 different images of people with different ethnicity and under various light conditions. The distribution of R-G is shown in Fig. 2, while that of R is shown in Fig. 3, and that of R/G is shown in Fig. 4. Our experiments revealed that 94.6% of the skin pixels have their R-G values between 20 and 90, see Fig. 2, which supports the observation in [9]. Also, we have noted that 96.9% of the skin pixels have their R value greater than 75, see Fig. 3, and 98.7% of them have their R/G values less than 2.5, see Fig. 4. We found out that adding a constraint on R and another on R/G improves segmentation.

It should be noted that this algorithm was able to perform with lower noise than the one proposed by Kovac et al, [10], even though it has less constraints. An example of this segmentation is shown in Fig. 5 and 6.

4 Face Localization

In this section, we will outline the steps we used in order to achieve the detection. Edge detection was used to separate the face from any non-face object.

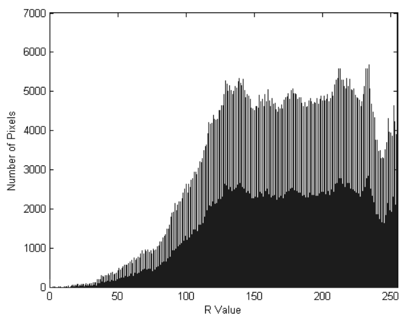

Figure 3: Distribution of R values in skin pixels

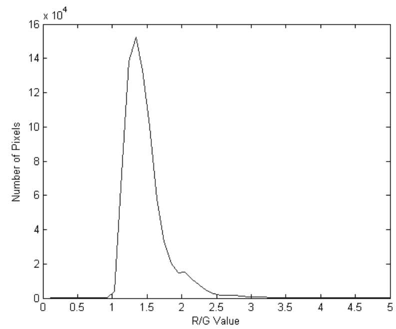

Figure 4: Distribution of R/G values in skin pixels

Figure 5: Original image

Figure 6: Segmented image

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 7: Sobel convolution masks

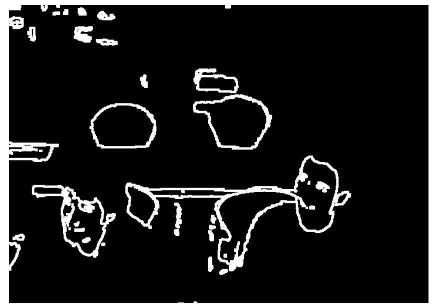

Figure 8: Image after applying edge detection and dilation

Later, the connected regions were grouped together and analyzed to determine which connected component is a face and which is not.

4.1 Edge Detection and Connected Components

As indicated earlier, edge detection is necessary to achieve a robust detection. A sobel edge detector, whose convolution masks are shown in Fig. 7, was used. The first mask estimates the gradient in the x-direction while the second estimates the gradient in the y-direction.

The gradient magnitude is calculated by:
$|G| = \sqrt{(G_x)^2 + (G_y)^2}$

An approximate magnitude which is faster to compute is given by:
$|G| = |G_x| + |G_y|$

The angle which gives rise to the spatial gradient is computed through:
$\Theta = arctan(Gy/Gx)$

After executing the edge detection, a dilation operation is applied to further separate the edges. The results of both processes are shown in Fig. 8.

The resulting image, combining skin color segmentation with edge detection, is inverted (to assign the “1” value to the blobs rather than the edges) and searched for connected components according to the adjacent 8-neighbor pixels.

4.2 Face Vs. Not-Face

Each of the connected components is then analyzed to judge whether it is a face or not. Simple shape features are used in the classification process. Each of these features, or cues, can be considered as a weak classifier. The cascade of these weak classifier forms a strong classifier, as demonstrated later in the experimental results section. These features are:

- **Area:** Normally, connected components with small areas correspond to noise generated by segmentation. Therefore, these regions are eliminated. Usually, any component whose area is less than 0.5% of the total area is dropped.

- **Bounding Box Properties:** The anatomy of the face suggests that the ratio of the bounding box height to its width is around 1.4 on average; however, this ratio varies slightly from one person to another and depending on the pose - whether it is frontal or profile. In our system, any region whose height is more than 1.9 times its width is removed.

- **Holes:** The face is a coarse surface in the sense that it has many curvatures. Therefore, it is expected to find a lot of holes in it when the edge detection is applied. The major three holes which are present in frontal face correspond to the two eyes and to the mouth; while in a frontal face we find one major hole corresponding to an eye.

- **Orientation:** There is a limit on how much me can "pan and tilt" our heads. Therefore, it is logical to expect that the orientation of any face blob, with respect to the x-axis, is – in absolute value – between 15° and 90° approximately. Any blob whose orientation is outside this range is dropped.

- **Centroid:** The face is evenly distributed in the region where it is located. Therefore, the centroid of a face region should be found in a small window centered in the middle of the bounding box. The dimensions of this window were found to be around 15% of the dimensions of the bounding box. Any region whose centroid is outside this window corresponds to a blob that is not evenly distributed and therefore it is not a face.

- **Extent:** The extent of a blob is defined as the area of this blob divided by the area of the bounding box surrounding it (both in pixels). Given the elliptical form of the face and its distribution, our experiments revealed that the extent for a face is between 0.45 and 0.75. Thus, any region whose extent is not in this range is eliminated.

5 Experimental Results

Many of the face databases commonly used by researchers include only gray-scale images, such as FERET face recognition database and the CMU face detection database [11]. Actually, the field of face detection lacks a database where the same agents are photographed under varying poses. Therefore, we have tested our method on images generated by a sequence we recorded. The scenario of this sequence includes one person entering a cafeteria, and then he is followed by two of his friends. The three of them sit together and chat for a while before they leave. The advantage of this sequence is that it provides us with faces having variations in pose, size, position, and expression.

We have collected 211 sample images, containing 266 faces, from our recorded sequence, and we have tested our method on. Out of the 266 faces, 238 were correctly detected with 22 false positives. The detection rate is 89.5% and the precision is 91.5% proving that our method is robust and efficient in detecting faces. Precision is defined as the ratio of detected faces to the sum of detected faces and false positives. The results are summarized in Table 1. Some of the tested images are shown in Fig. 9. Please note the variations in pose, scale, position, and expression. Obtaining such a high detection rate on images under such variations is considered a good result.

Number of Faces	266
Positive Detections	238
Detection Rate	89.45%
False Positives	22
Precision	91.5%

Table 1: Results Summary

Figure 9: Experimental Results

6 Conclusions

In this paper, a novel approach for face detection in color images is proposed. Skin color segmentation is applied to separate the skin areas from the non-skin. Edge detection with dilation is then implemented to separate face candidates from any background or foreground blob. Connected components are later analyzed using primitive shape features to decide which blob is a face and which is not. The experimental results revealed the robustness and efficiency of this method under varying conditions.

Future work includes adaptive face detection and tracking based on active cameras. The work will address the problem of facial features detection and tracking in real-time using a single active camera. The variable parameters of the camera (i.e. pan, tilt, and zoom) are changed adaptively to track the face of a single agent in successive frames. The performance of this approach should be independent of the velocity of the agent, and is robust even to partial occlusions. A multi-zoom framework for activity analysis will be investigated for situations requiring combinations of both detailed and coarse views of the scene.

Acknowledgements

This work is supported by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDI-video project, and by the Spanish MEC under projects TIN2006-14606 and DPI-2004-5414. Jordi Gonzàlez also acknowledges the support of a Juan de la Cierva Postdoctoral fellowship from the Spanish MEC. The authors thank Dr. Ramon Baldrich for providing valuable assistance to this paper.

References

[1] Yang, J., Kriegman, D.J., and Ahuja, N., “Detecting Faces in Images: A Survey”, *IEEE Trasactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 1, pp. 34-58, 2002.

[2] Viola, P., and Jones, M., “Robust Real-time Object Detection”, *Proceedings of IEEE Workshop on Statistical and Computational Theories of Vision,* Vancouver, Canada, July 2001.

[3] Huang, C., Ai, H., Wu, B., and Lao, S., "Boosting Nested Cascade Detector for Multi-View Face Detection", *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04),* vol. 2, pp. 415-418, August 2004.

[4] Nara, Y., Yang, J., and Suematsu, Y., "Face Detection Using the Shape of Face with Both Color and Edge", *Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems,* vol. 1, pp. 147-152, December 2004.

[5] Ravi Kumar, C.N., and Bindu, A., "An Efficient Skin Illumination Compensation Model for Efficient Face Detection", *Proceedings of IEEE 32nd Annual Conference on Industrial Electronics,* pp. 3444-3449, November 2006.

[6] Hsu, R.L., Abdel-Mottaleb, M., and Jain, A.K., "Face Detection in Color Images", *IEEE Trasactions on Pattern Analysis and Machine Intelligence (PAMI),* vol. 24, no. 5, pp. 696-706, 2002.

[7] Vezhnevets, V., Vassili, S., and Andreeva, A., "A Survey on Pixel-Based Skin Color Detection Techniques", *Proceedings of International Conference on Computer Graphics (GRAPHICON'03),* Moscow, Russia, September 2003.

[8] Shin, M.C., Chang, K., and Tsap, L.V., "Does Colorspace Transformation Make Any Difference on Skin Detection?", *Proceedings of 6th IEEE Workshop on Applications of Computer Vision (WACV'02),* pp. 275-279, December 2002.

[9] Al-Shehri, S.A., "A Simple and Novel Method for Skin Detection and Face Locating and Tracking", *Lecture Notes in Computer Science, Computer Human Interaction,* vol. 3101, pp. 1-8, 2004.

[10] Kovac, J., Peer, P., and Solina, F., "Human skin color clustering for face detection", *Proceedings of IEEE Region 8 Computer as a Tool (EUROCON'02),* vol. 2, pp. 144-148, September 2003.

[11] Zhang, H., and Zhao, D., "Spatial Histogram Features for Face Detection in Color Images", *Lecture Notes in Computer Science, Advances in Multimedia Information Processing (PCM'04),* vol. 3331, pp. 377-384, 2004.