IMPLEMENTATION OF FAST FACE DETECTION ALGORITHMS FOR MOBILE PHONES

Sandro da Silva
Department of Computer Science
University of the Western Cape
Bellville 7535, Cape Town
South Africa
sdasilva@uwc.ac.za

Johnson I. Agbinya
Faculty of Engineering (I&C Group)
University of Technology, 1 Broadway
Sydney NSW 2007,
Australia
agbinya@eng.uts.edu.au

Abstract

Network access control using face recognition increases the user-friendliness in mobile phones and human-computer interaction. In order to realize a real time system implemented on handheld devices with low computing power, low complexity algorithms for face detection are required. This paper introduces a fast algorithm for face detection in video sequences, using a skin color model in the HSV color space. These algorithms are fast, robust and reliable and they permit the segmentation and tracking of multiple instances of faces in thumbnail images. The method is designed and implemented to work well with color images with varying lighting conditions and inhomogeneous background. The paper targets mobile handsets such as the NOKIA 9500 Communicator.

Keywords: face detection, RGB, HSV, 3G networks, mobile phone handsets, wireless devices.

1. Introduction

The human face provides a variety of different communicative functions such as identification and the perception of emotional expressions. Biometric systems using face recognition are attracting attention for authentication and authorisation. Similarly, the cost effectiveness of sending only a facial image over a narrowband wireless network instead of a full size photo can hardly be contended. In order to realize a real time system on handheld devices with low computing power, we need to implement low complexity algorithms for face tracking. Using skin color for the detection of human faces has been proved to have several advantages compared to other well-known methods, because processing for color information is much faster than processing other facial features and under constant lighting conditions color is nearly invariant against changes in size, orientation and partial occlusions of the face [1]. This paper addresses face detection based on color techniques. The rest of this paper describes the state of the art. Then we discuss the proposed face detection technique using a skin color model in the HSV color space. Finally, we present the experimental results.

2. Related Research

The goal of face detection is to find a face in the image and, if found, return the location of the image and the extent to each face. Some of the face detection applications can be part of face recognition, a surveillance system, or video based computer machine interface. Efficient face detection at frame rate on mobile phones is an impressive goal; it is analogue to face tracking that requires no knowledge of previous frames. Fast face detection can be used to initialize face tracking.

Many methods have been proposed to build a skin color model. The simplest model is to define a region of skin tone pixels using Cr, Cb values [2], i.e., R (Cr, Cb), from samples of skin color pixels. With chosen thresholds, $[Cr_1, Cr_2]$ and $[Cb_1, Cb_2]$, a pixel is classified to have skin tone if its values (Cr, Cb) falls within the ranges, i.e., $Cr_1 \leq Cr \leq Cr_2$ and $Cb_1 \leq Cb \leq Cb_2$. Kjeldsen and Kender defined a color predicate in HSV color space to separate skin regions from background [6]. Gaussian density functions [7] and a mixture of Gaussians [8] are often used to model skin color. The parameters in a unimodal Gaussian distribution are often estimated using maximum-likelihood [9].

These methods were all derived for PC applications. Applications in handsets require less complexity but efficient computing. That is the objective of this paper.

3. Face Detection

3.1 HSV Color Model

We use skin color as the basic feature for face detection. To segment human skin regions from an image, a reliable skin colour model that is adaptable to different skin colours and to different lighting conditions is needed. The basic RGB colour model is not suitable for skin colour characterization because the triplet (R, G, B) represents not only colour but also luminance, which varies across the person's face due to the ambient lighting. In mobile phone photos, the face often dominates the photo in terms of the number of pixels used to represent it compared to the other parts of the body. Therefore, algorithms using face colour features are likely to perform better on handsets. One of the most significant parts of this project was to find an appropriate skin color model in order to facilitate real-time face detection, which is adaptable to people of different skin colors and to different lighting conditions. Thus, we have found a skin color model in the HSV color space to be appropriate, because HSV gives the best performance for skin pixel detection [10].
Most colour pictures are recorded as (R, G, B) triplets. Given a colour defined by (R, G, B) where R, G, and B are normalized to 0.0 to 1.0, an equivalent (H, S, V) colour is determined by the following set of formulas (1) and (2). Considering MAX to be the maximum of the (R, G, B) values, and MIN the minimum of those values, the formula is then expressed as [11]:

$$H = \begin{cases} 0 + \frac{G-B}{MAX-MIN} x60, \text{ if } R = MAX \\ 2 + \frac{B-R}{MAX-MIN} x60, \text{ if } G = MAX \\ 4 + \frac{R-G}{MAX-MIN} x60, \text{ if } B = MAX \end{cases} \quad (1)$$

$$S = \frac{MAX-MIN}{MAX}; \quad V = MAX \quad (2)$$

Where H varies from 0.0 to 360, indicating the angle in degrees, S and V vary from 0.0 to 1.0.

3.2 The Proposed Face Detection Method

First of all, we create a skin color model in the HSV color space. To create the skin color model, we cut skin regions from 23 true color images of people of different ethnic backgrounds. Each of these skin color samples is then converted to the HSV color space. The conversion of each color image to HSV space eliminates the luminance component of the RGB space, because lighting effects change the appearance of the skin. The color histogram revealed that the color distributions of skin color of different people are clustered in the H-S color space.
We then build a mean and covariance model of H- and S-values from the HSV[1] color space:

$$Mean: m = E\{x\} \quad (3)$$

$$Covariance: C = E\{(x-m)(x-m)^T\} \quad (4)$$

Where, $x = (H\,S)^T$. With the mean and covariance values, the skin color model is fit into a Gaussian model by N(m, C), by plotting H against S.
With the skin color model, the segmentation process can begin. The first step is to modify the original image in which we want to detect a face. Thus, to eliminate the luminance component of the original image, in the RGB color space, we convert the image into the HSV color space as indicated previously.
Then we segment the original image to find regions that are most likely a skin region using the skin Gaussian model and the following expression:

$$P(H,S) = \exp[-0.5(x-m)^T C^{-1}(x-m)] \quad (5)$$

Where $x = (H\,S)^T$
The above expression determines the probability of one pixel being a skin region based on the skin color model. This probability is given on a grayscale skin probability image based on the gray level.
Further segmentation is then done to threshold the grayscale skin likelihood image into a binary image. Since skin colors vary between people, an adaptive threshold process to find the optimal threshold value is used. If the threshold value is too low, the amount of segmented skin regions increases. Based on this assumption, an adaptive threshold is created. This method starts the threshold value at 0.65 and decrements by factors of 0.1 until it gets to 0.05. The program determines the optimal threshold value by finding the point when the change in the number of segmented regions is minimum. After the optimal threshold value is determined, the grayscale image is then converted to binary. In

[1] The values for V were discarded, only took into consideration the H- and S- values.

the binary image, the white regions show the skin regions and are labeled as 1 and the black regions show the non-skin regions and are labeled as 0. Each of these skin regions is assigned an integer value, and they are tested individually to see if it represents a face.

Furthermore, we proceed to determine which regions can possibly determine a human face. We look for a closed white (skin) region that has one or more holes (black region) inside it. The black holes inside white regions result from facial features such as nose, mouth, and eyes. Thus, to determine the number of holes in a white region we use the following expression:

$$H = 1 - E \qquad (6)$$

Where H is the number of holes in a region and E is the Euler number. We use 1 because we are analyzing only one segmented region at a time.

If an area with at least one hole in it is found, we continue to find other characteristics about the region before concluding that this area constitutes a face.

In order to include faces with an inclination, with the intention of maximizing the detection rate, we compute the angle of inclination of the face.

After determining the center and the angle of the white region with at least one hole in it, we now calculate the width and the height of that region to improve the decision process. This includes moving one pointer from the left, one from the right, top and bottom of the image. We compute the coordinate of 4 boundary points, and calculate the height by subtracting the bottom and top values and the width by subtracting the right and the left values.

Another parameter that is also useful to determine if the segmented area is a face is the ratio of height to width. The height to width ratio of a human face is normally 1. In order to improve the detection rate, we found experimentally that 0.8 is a good minimum value. Values below 0.8 are not helpful to conclude that the region is a face, because human faces are oriented vertically. We determined a good upper limit to be 1.9. However, there are situations where we have indeed human faces but the upper limit of the ratio is higher. This happens when the neck of the person is uncovered. Therefore, to account for this we set the ratio to be 1.9 and we eliminate the region below.

A 2-Dimensional template matching algorithm was used to build a template face which is used to take the final decision of determining if the skin region represents a face. The template face grasps as much as possible the common features of human face, but is not dependent on the background and individual characteristics of the face. The average face template is generated by enclosing the eye brows and the upper lips.

The template face image is used to fill the area of the image corresponding to the skin region. This process is done by resizing and positioning the template face according to the skin region's characteristics. Thus, the width and height values are used to resize the frontal face model into these dimensions. The value of the angle of inclination is used to rotate the resized template face into the same direction as the skin region. The center of mass is used to place the template face exactly at the center of the skin segmented region.

A process using two dimensional correlation coefficients between two matrices is used to determine how well the template face fits into the skin region. This function is the implementation of the following algorithm:

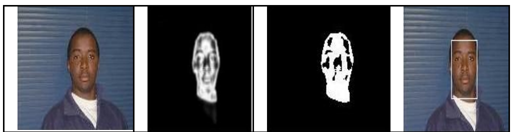
$$r = \frac{\sum_m \sum_n \left(A_m - \overline{A}\right)\left(B_m - \overline{B}\right)}{\left[\sum_m \sum_n \left(A_m - \overline{A}\right)^2\right]\left[\sum_m \sum_n \left(B_m - \overline{B}\right)^2\right]} \qquad (7)$$

Where A_m and B_m are matrices of the same size, $\overline{A} = means2(A_m)$, and $\overline{B} = means2(B_m)$.

The return value r is scalar double.

A good threshold value for classification of a skin region as a face is if the resulting correlation value is greater than 0.6.

Once every region has been successfully evaluated, the original color image is displayed (**figure 1**) with rectangles placed around the faces in the image. This is done by obtaining the coordinates of the part of the image that has the template face (model). With this coordinates, we draw a rectangle in the original color image.



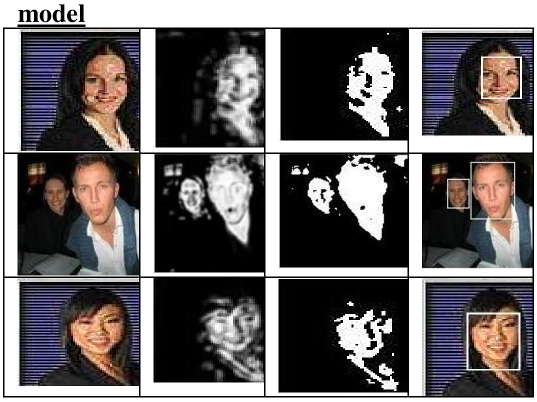
Original Image	Skin-likelihood image	Skin-segmented image	Final result

Figure 1: General Overview of the described method

4. Experimental Results

The system was tested on a set of 51 images, of who Set 1 are Caucasians, Set 2 are Africans, and Set 3 are Asians, both males and females. This selection of various ethnic groups in the test set was to ensure the robustness of the system. All the images include varying illumination conditions and varying background conditions. **Figure 3** illustrates the system's face detection capability using the HSV color model, from (a) to (d), the first images are the original test images that are being tested for instances of a face, next are the skin likelihood images, then are the segmented binary images, the last images are the original images with a box drawn around the face. The program worked quite well in example 1 of **figure 3**, where in the image we have a face of dark skin color. The system successfully recognized the single instance of a face in the image, but it also included the area of the neck. This happened during the labeling of each area, the neck was included with the face area as one labeled region. In example 2, where we have an image of a Caucasian lady, the program successfully detected the single instance of the face in the image.

Figure 2: Tests using the HSV Skin Color model



(a) **(b)** **(c)** **(d)**

Figure 2: (a) The original color images; (b) Skin-likelihood images; (c) skin-segmented images; (d) final results

For example 3 of **figure 2**, it is a demonstration of how multiple instances of faces can be detected. We have in the image a Caucasian couple, and the program successfully recognized the two instances of the faces in the image even when the face at the background is highly faded. In the last example we have an image of an Asian (Chinese) lady with her hair covering part of her forehead. The program has worked relatively well and the hair in her forehead did not affect the detection results. However, part of the area of her neck was included in the face region. This obviously happened during the labeling process as described for example 1. Otherwise, the system successfully recognized the single instance of a face in the image.

Table 1 gives the face detection results for the HSV technique based on 66 faces. Of these, 59 faces are correctly detected, 7 are missed and 2 faces are falsely detected. The likely reasons for the faces that are missed result from bad lighting conditions and background distractors.

Set of color Images	Correctly detected	Missed	Falsely detected
Set 1	25	3	0
Set 2	19	2	0
Set 3	15	2	2
Total	59	7	2

Table 1. Performance of algorithm with HSV model

Another reason for the undetected faces, is due to the different color of skin or face profiles found across several subjects.

We found experimentally that it is possible to separate human skin regions from complex background based on the HSV color space, because its color map is the most adequate for differentiating the skin regions from the rest of the photo contents. By plotting the skin regions vs. non-skin regions in a graph (**figure 3**) of H vs. S we can maximize the amount of skin pixels and minimize the number of background pixels by setting thresholds.

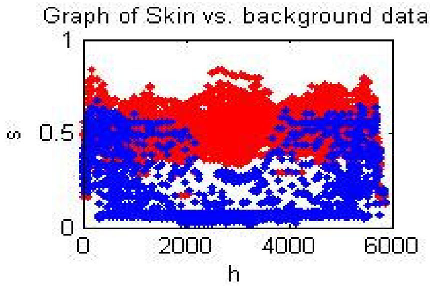


Figure 3: Skin (blue) vs. background (red) in HS space

From the above plot of a test image we observed that there is less overlapping of the non-skin pixels with the skin pixels in HS space.
The proposed method is not totally accurate; therefore, it was necessary to run a set of tests to measure the accuracy of the program. During the testing we encountered that the optimal threshold values varied from 0.1 to 0.5. The optimal threshold values with more detection of faces were 0.5 followed by 0.4. We also observed experimentally that minimum value for the height to width ratio was about 0.8, and the maximum value was around 1.9. Cases with ratio below 0.8 were usually a miss, and with values above 1.9 few instances of faces were detected but they included large parts of the neck and the hair. For the height to width ratio on the range from 0.8 through to 1.9 the least amount of non-faces is found and the highest amount of faces was detected. Most of the non-face regions were eliminated by the height to width ratio test.
Through these tests we found out that the best correlation value was about 0.6, because the highest percentage of faces is detected and the least amount of non-faces is found at this value. Moreover, the correlation values incremented from 0.1 to 0.8.
We ran tests on 51 distinct images with a total of 66 faces in order to determine the percentage of faces detected by our approach. With the best value of 0.6 for the correlation and the ratio of height to width on the range from 0.8 to 1.9, the percentage of faces included on the rectangle of the output image was around 89 % (59 faces). The face is considered detected if it is inside the final result box. We encountered one image that included two faces in a box, but counted these two faces as being detected. In some images the final result box only contained part of the persons face, for example the chin is cut out but the other facial features are included.
To measure the accuracy of the program we noticed that in very few cases non-face areas were also detected. To be precise, in two instances some non-face areas were falsely detected as faces. It is cumbersome to give results regarding the non-face images because these results were reliant on the specific images. However, it is necessary to have a high percentage of faces detected with non-skin areas occasionally being detected rather than faces in images not being detected.

4.1 Computing Time

Our data shows that the runtime for this approach scales linearly with the number of people in the test image (for N people, runtime ≈ (N + 1) x 200ms). For real-time performance, a face detection domain is limited to 1 person on an image with face size between 40 x 40 to 70 x 70 pixels (around 0.4 seconds) when running the detection code on handheld computers.
The memory footprint of the face detection algorithm is (3 + N x 0.006) MB for N people. This scales to just over 3MB for 20 people, which is reasonable for handheld devices.
The runtime during the training phase for our 23 skin samples to build the skin color model is around 5 seconds. And we did not need to train non-skin samples, yet the performance is 89 %. And the memory footprint for the training phase of the 23 skin samples to build the skin color model is 3.138MB.
The memory footprint scales to just over 3MB for 20 people during the operational portion and 3.138 MB to build the skin color model of 23 skin samples, which is reasonable for handheld devices.
The runtime to build the skin color model is around 5 seconds, and the real-time performance when running the detection code on a handheld is around 0.4 seconds per image.
This application targets mobile handsets such as the NOKIA 9500 Communicator. The Nokia emulator can be integrated with Sun One Studio and Nokia Developers Suite. The emulator allows us to test the application as if it is being tested on an actual mobile handset.

5. Conclusions

We have proposed and implemented a fast and reliable algorithm for face detection based on the HSV-colour space, with adaptive threshold which simplify the detection of faces within video sequences on mobile phones and that are suitable for implementation on mobile handsets. The algorithm is used to segment and detect multiple faces in images. We observed experimentally that it is possible to separate human skin regions from complex background using HSV color space. Moreover, the HSV color map is suitable for differentiating the skin regions from the contents of the rest of the image, because there is less overlapping of the non-skin pixels with the skin pixels, if you make a plot of the skin pixels vs. the background in the HSV space. The HSV color space is also adequate to get rid of the undesirable skin-color-like pixels of the background.

The percentage of faces detected by our approach using the skin color model in the HSV color space was around 89 %, using the best value for the correlation of 0.6 and the height to width ratio being on the range 0.8 to 1.9.

This face detection technique, besides being used as a prior step for face recognition using handheld devices in remote surveillance, it could also be used in a wide range of applications such as a human motion detection system to avoid false alarms in secure areas, gesture recognition to bridge the digital divide that exists with the deaf community, security control, detection and tracking of online pornographic images, video retrieving, image database management, biometric signal processing, human computer interface, face recognition.

ACKNOWLEDGEMENT

Sandro da Silva thanks the Telkom Center of Excellence at the University of the Western Cape for their financial support.

REFERENCES

[1] A. Albiol, L. Torres, C.A. Bouman, and E.J. Delp, "A simple and efficient face detection algorithm for video database applications," in Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, September 2000, vol. 2, pp. 239-242.

[2] D. Chai and K. N. Ngan, "Locating Facial Region of a Head-and-Shoulders Color Image," Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, 1998, pp. 124-129.

[3] J.L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1997, pp. 640-645.

[4] D. Saxe and R. Foulds, "Towards Robust Skin Identification in Video Images," Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, 1996, pp. 379-384.

[5] M.J. Swain and D.H. Ballard, "Color Indexing," Int'l J. Computer Vision, vol. 7, no.1, 1991, pp. 11-32.

[6] R. Kjeldsen and J. Kender, "Finding Skin in Color Images," Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, 1996, pp. 312-317.

[7] J. Cai, A. Goshtasby, and C. Yu, "Detecting Human Faces in Color Images," Proc. 1998 Int'l Workshop Multi-Media Database Management Systems, pp. 124-131, 1998.

[8] M.-H. Yang and N. Ahuja, "Gaussian Mixture Model for Human Skin Color and Its Application in Image and Video Databases," Proc. SPIE: Storage and Retrieval for Image and Video Databases VII, vol. 3656, 1999, pp.458-466.

[9] S.-H. Kim, N.-K. Kim, S.C. Ahn, and H.-G. Kim, "Object Oriented Face Detection Using Range and Color Information," Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, 1998, pp. 76-81.

[10] B. D. Zarit, B. J. Super, and F. K. H. Quek, "Comparison of five color models in skin pixel classification," In ICCV'99 Int'l Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems, 1999.

[11] Da Silva, S., and Agbinya, J.I., "Face Recognition Programming on Mobile Handsets", To appear in the Proceedings of ICT 2005 - 12th International Conference on Telecommunications, Cape Town, South Africa, 3 – 6 May 2005.

Authors Biography

Sandro da Silva: obtained B.Sc. honors in computer science in 2002 at the University of the Western Cape. He is currently a M.Sc. candidate. His research interests are: Computer vision, digital image processing, human-computer interaction, mobile networks.