Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Improved Automatic Skin Detection in Color Images

Filipe Tomaz and Tiago Candeias and Hamid Shahbazkia

Universidade do Algarve, FCT
Campus de Gambelas - 8000 Faro, Portugal,
{ftomaz, tcandeia, hshah} @ ualg.pt
Home page: http://w3.ualg.pt/∼ftomaz/fr/fr.php

Abstract. Mahalanobis distance has already proved its strength in human skin detection using a set of skin values. We present this work that uses automatic skin detection after an initial camera calibration. The calibration is done by human sampling from test individuals. A scaling is performed on the work data, before applying the Mahalanobis distance that ensures better results than previous works. We use the TSL color space also used successfully by others authors, where undesired effects are reduced and the skin distribution fits better in a Gaussian model than in others color spaces. Also, using an initial filter, normally large areas of easily distinct non skin pixels, are eliminated from further processing. Analyzing and grouping the resulting elements from the discriminator, improves the ratio of correct detection and reduce the small non skin areas present in a common complex image background, including Asiatic, Caucasian, African and interracial descent persons. Also this method is not restricted to orientation, size or grouping candidates.
The present work is a first step in a approach for human face detection in color images, but not limited in any way to this goal.

1 Introduction

The need for better human computer interaction is now a fact and probably will be in the next few years. More friendly and effective methods regarding human activity are in constant development to free users from manual interaction. Gesture recognition, robot interaction, multimedia, face detection and recognition, hand detection, teleconference and many other applications are, or can be based on skin detection to restrict the complexity of further processing. Size, quality and cost of image acquisition are also very important to the rapid growing needs for such systems and commercial applications are now available. We propose a method that can handle a wide range of variations in skin colors [6]. Further more, since we make a pixel based segmentation, size and orientation are irrelevant. We also intend to apply this work without constraining the background.
In this paper, we describe a general, robust skin detection method based on a scaling of the calibration inputs discussed in section 2. Next, a filter is applied on the working image that reduces the computational costs of the segmentation

method covered in section 4. In section 5 a threshold is used to discriminate the skin from non skin pixels that will be grouped in larger objects, described in section 6. Some results are presented and a hopefully discussion is present in section 8 where the relevant aspects, and future research directions of the implementation are focused.

2 Calibration and scaling

Calibration of the input device is done taking samples of different skin colors and locations. Terrillon et al. [2] have suggested a TSL (tint saturation luminance) color space in its comparative performance evaluation. As explained in section 4 we need to retrieve from the samples a saturation (S) and a tint (T) value for each sample pixel that are normalized in the range [0.; 1.0] (reference [1]) before applying the Mahalanobis distance.
Since Asians and Caucasians have very similar skin color, often a skin chrominance distribution is modeled by only one Gaussian [2, 1]. A more general approach (more skin colors) need a more complex model, that will be certainly more computational demanding. Several studies already showed, that the major difference in the skin color appearance lies, not in color itself but in the intensity [4]. Departing from this argument we studied and tested a new scaling method for the testing pixels based on the calibration values that improved the overall results. Using a simple normalization for the saturation value in [0.; 1.0] (equation 1) range, and a scaling for the tint value obtained by eq. 2, based on the saturation calibration minimal value, we obtained good results as shown in section 7, for white, black skin and descents, only using one skin chrominance model.

$$S = (S - MinS)/(MaxS - MinS) \tag{1}$$

$$T = (T - MinS)/(MaxT - MinS) \tag{2}$$

Figure 1 shows the ST normalized color space, and with the scaling that we used. Those values were obtained, using 26 samples from different persons and different skin colors. For example purposes the surrounding region of each ST point is highlighted. As shown, in the plot with the tint scaling, the ST region for skin is located above than the ST normalized space. This means that we consider the tint value more discriminative than using the normalized ST value in the equation 8 presented in section 4.

3 Initial filter

In any image with a complex background, the skin area is often smaller than the non skin pixels. Since we use a pixel evaluation procedure, to reduce the computation effort, we apply a initial filter that can remove all the pixels easily labeled as non skin. This filter only aims to remove pixels from further processing,

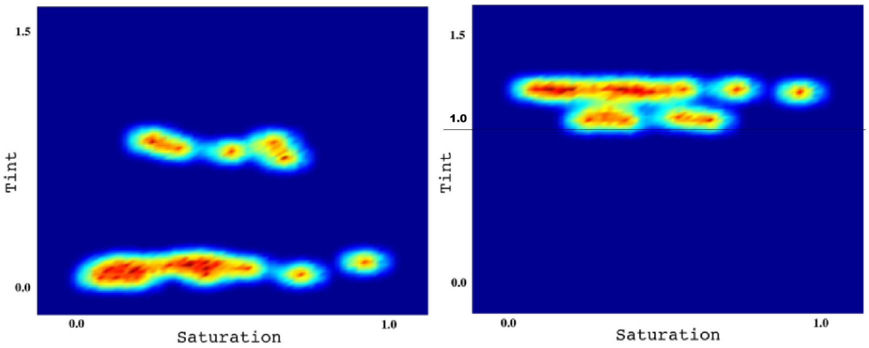

Fig. 1. ST space (Left: ST space normalized. Right: with tint scaling used)

and hopefully skin pixels are not removed. Green, blue, yellow and other well defined colors are expelled using empirical rules defined in the code presented in fig 2.

```
// R G and B values are presented in the range 0 to 255

   if ( (B > 160 && R < 180 && G < 180) || // Too much blue
        (G > 160 && R < 180 && B < 180) || // Too much green
        (B < 100 && R < 100 && G < 100) || // Too dark
        (G > 200) || // Green
        (R+G > 400) || // Too much red and green (yellow like color)
        (G > 150 && B < 90 ) || // Yellow like also
        (B/(R+G+B) > .40) || // Too much blue in contrast to others
        (G/(R+G+B) > .40) || // Too much green in contrast to others
        (R < 102 && G > 100 && B > 110 && G < 140 && B < 160) // Ocean
)
```

Fig. 2. Source code that implements the initial filter.

The presented values are well suited for our application, and tested with three different input devices, but also can be easily changed for any purpose, or if skin pixels are removed.

4 Skin segmentation

From Terrillon [2] previous work we can observe that the TSL chrominance space is very effective for skin segmentation when using a Gaussian model. This is also true where illumination conditions vary. Using the following transformations the equivalent pixel representation is obtained on the ST color space.

$$S = \sqrt{\frac{9}{5\left(r'^2 + g'^2\right)}} \qquad and \qquad T = \begin{cases} \frac{tan^{-1}\left(\frac{r'}{g'}\right)}{\pi} + 0.5 & g' \neq 0 \\ 0 \quad g' = 0 \end{cases} \tag{3}$$

where,

$$r' = \frac{R}{R+G+B} - 1/3 \qquad and \qquad g' = \frac{G}{R+G+B} - 1/3 \tag{4}$$

Defining the class Cs that represents skin color as equation 5,

$$C_s = \begin{bmatrix} \sigma^2 M_t & \sigma T M_s \\ \sigma T M_s & \sigma^2 M_s \end{bmatrix} \tag{5}$$

and using the Mahalanobis distance, from the mean vector v_m defined by equation 6.

$$v_m = [M_t \quad M_s]^T \tag{6}$$

the Mahalanobis distance obtained by equation 7, where $\lambda_{i,j}$ is the distance of pixel$_{(i,j)}$ to the v_m.

$$|\lambda_{i,j}^2| = [X_{i,j} - v_m]^T C_s^{-1} [X_{i,j} - v_m] \tag{7}$$

that result in equation 8.

$$\lambda_{i,j}^2 = \left(\frac{T_{i,j} - M_t}{\sigma^2 M_t T_{i,j}^2 M_s} - \frac{S_{i,j} - M_s}{\sigma T_{i,j}(\sigma^2 M_t - T_{i,j}^2 M_s)} \right) (T_{i,j} - M_t) + \left(\frac{-(T_{i,j} - M_t)}{\sigma T_{i,j}(\sigma^2 M_t - T_{i,j}^2 M_s)} + \frac{(S_{i,j} - M_s) M_t}{M_s(\sigma^2 M_t - T_{i,j}^2 M_s)} \right) (S_{i,j} - M_s) \tag{8}$$

5 Threshold

While Terrillon et al. [1] uses a fixed threshold, we adopted a different method that analyse all the image before a binary selection. This method is better for illumination variations and for different skin colors. Calculation is done using the $\lambda_{i,j}^2$ value obtained by equation 8. Since this value represents the Mahalanobis distance to the mean vector of the initial samples v_m, we then normalize $1 - \lambda_{i,j}^2/cons$ to the range [0.: 1.0]. Only after this normalization we threshold the pixels as skin, if the value is greatest than 0.7.

The *cons* value is obtained by direct experiment and should be tuned for different input devices.

This threshold method, always select part of an image to the maximum value 1 and probably, if similar pixels are present, values very close to 1 will also be present. This can be seen as a solution to maximize the difference among candidates or a problem if no skin is present, since at least a non skin pixel will match 100% as being skin, in result of the normalization. In our point of view, and also from direct experiment, images where skin do not exist, do not have a large group of pixels similar to skin, and normally they are isolated. In this way the grouping method described in section 6, can manage this problem. Also, if skin is present, is expected that the $\lambda^2_{i,j}$ will be smaller for skin pixels. Even if skin is very different from the calibration data used, $\lambda^2_{i,j}$ is expected to be smaller than those that are not skin, so the skin pixels will also tend to approximate the value 1.

6 Grouping

The resulting image from the previous section can be enhanced, also if we are looking for skin regions, a grouping method should be used. After the image threshold we apply a median filter to smooth the image, and to reduce the small amounts of isolated noise resulting from the method descrived in section 4. Only small areas are modified and almost no skin pixels are erased.

Using a connect pixels analysis, we obtain groups **G** that are initially filtered by their size, if they present a small number of pixels in relation to the image size. We also connect from the initial universe **G**, two or more close groups to reduce some common errors (as example, two fingers separated). This was done by finding the smaller distance between groups that must be proportional to the size of one of the elements.

Using the image size, we do not need a human interaction, and tests showed the good results for images from size $320x200$ up to $1600x1200$. In the figures group (fig. 3) we can observe that although separated, the skin elements are grouped and are returned as a single one.

This grouping method is useful for face detection and hand detection. Close elements from distinct objects are also returned connected if the grouping evaluation is satisfied, so the grouping should always be performed with caution.

7 Results

Using only the tint normalization of the input data, resulted in a better quality pre-threshold image than previous works [1], and the skin color range is also improved. We tested the implementation in different images with simple and complex backgrounds, indoor and outdoor, image sizes and skin colors, from Caucasians, Africans descents and Asians. The results are good since we have a robust and noisy free algorithm that can suit almost any kind of system were skin detection is important. To ensure better results, a calibration should be performed in the camera were future input images will be obtained, although not necessary.

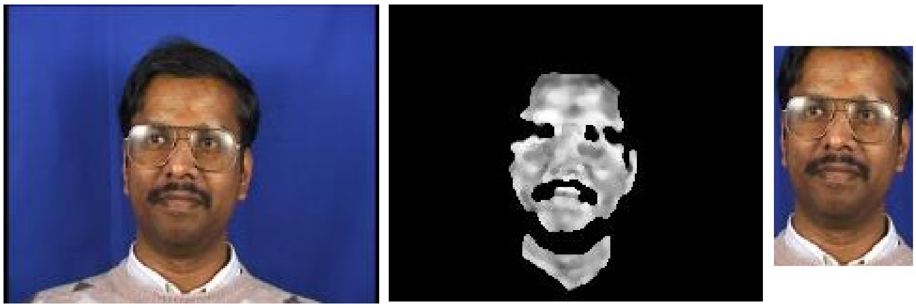

Fig. 3. Grouping example. Although the neck is separated from the face, the two components are still treated as a single group. From left to right: Original image; Median image that resulted from $\lambda_{i,j}^2$ calculations; minimal rectangular region that contained the pixels labeled as skin.

Using a Philips ToUCam, we obtained 50 test images containing 78 persons, 42 white Caucasians, 12 dark skin, 2 Asian descents and 12 white-dark skin descents, in an environment with small light variations. We collected about 8 pixels for each person, from locations that we understood as important, mainly face (around the eyes) arms and neck.

As stated before, the *const* value presented in section 5 was obtained by inspection, were we attempted to reduce the amount of non skin pixels labeled as skin, since our goal is to deliver free skin areas to further processing. Reporting the exact percentage of correct labeled pixels is a difficult task since we ca not manual label, at the pixel level, all the tested images.

Since we based this work on color segmentation, the system is also susceptible to mistakes if the background, clothing or other objects are present with colors that fit in the range defined as skin by the presented algorithm. This is of course the greatest limitation of a color pixel based method to skin detection.

Unfortunately there are not, to our knowledge, any color image database that could be used as a comparison to previous works. In the figures groups labeled 4 to 8 we demonstrate the system ability's.

8 Discussion

As previous mentioned, better results are obtained when the calibration is done with the input device. Although true, the system also behaviors very well if the image is obtained by another device. Using a Sony P51 camera, without any changes in the calibration data obtained in a Philips ToUCam, we also obtain clean skin regions, even when the light conditions vary (for example, image 4 and 6).

The robustness is in fact one of the main characteristics of this color segmentation, but vulnerable when objects are present with a color similar to skin. Only a high observation level can manage this problem. High light variations

Fig. 4. Very complex picture with different sizes groups of skin. Showing the original image, the median image and the group resulting analysis. Only one group is not correctly classified since it have a very similar color to skin (food in table). The grouping algorithm also connects two separated elements (forehead from the rest of the face) that compose the man with glasses. **Image size**: 800x600. **Input device**: Sony P51. **Comments**: Night picture with flash, indoor.

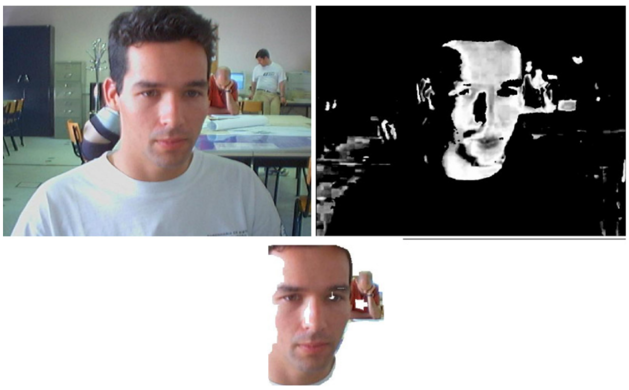

Fig. 5. Complex picture with frontal face and high light variation. Showing the original image, the median image and the group resulting analysis. The effect of light difficult the correct extraction of the face. **Image size**: 640x480. **Input device**: Philips ToU-Cam. **Comments**: Daylight, indoor.

(see image 5 and 8) reduces the chromatic information, that result in a small detection ratio, the image 5, shows part of a face affected by intense daylight that difficult the final result. A light compensation similar to the one presented in [3] could be used. Using a dinamic color model also could improve the detection of skin where light sources vary [5], mostly in video images.

Depending on the goal of each application, the value *const* can be changed. Since we use samples of different skin origins, we can also improve the results if we know in advance the type of input images that the system will have to manage. We used 42 Caucasians, 12 dark skin, 2 Asians descents and 22 white-

Fig. 6. Small complex background with frontal different color skin faces. Showing the original image, the median image and the group resulting analysis. One white Caucasian and two white-dark skin descents are present. The neck was discarded because it possesses a small number of pixels. **Image size**: 800x600. **Input device**: Sony P51. **Comments**: Daylight with artificial light, indoor.

dark color descents, because the average population of the University of Algarve probably have a similar distribution.

Some work will be addressed to solve some of the problems mentioned before, although the system behaviors very well to match almost any application requirements.

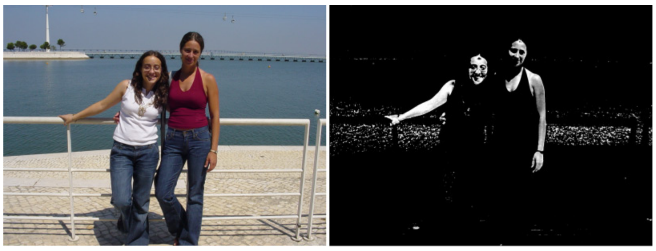

Fig. 7. Simple background. Showing the original image and the median image. **Image size**: 1600x1200. **Input device**: Sony P51. **Comments**: Daylight, outdoor.

Notes and Comments: This prototype was implemented in a Linux OS, Mandrake 9.1 with kernel version 2.4, using the following tools among others: QT 3.1.2 from Trolltech, ImageMagick-5.5.6 and GCC 3.2.2. As input devices, a Philips ToUCam Pro, a Sony P51 and a SonyP31 were used.

Fig. 8. Simple background with different color skin arms and hands. Showing the original image and the median image. The effect of light is observed in the bigger hand and on the arm present on the right top corner. **Image size**: 1600x1200. **Input device**: Sony P51. **Comments**: Night with artificial light and flash, indoor. Cheers !

References

1. Terrillon J-C., David M., and Akamatsu S.: *Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments.* In Proc. of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998. pp. 112-117 url = "citeseer.nj.nec.com/terrillon98automatic.html"
2. Terrillon J-C., Shirazi M., Fukamachi H, and Akamatsu S.: *Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images.* Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 54-61, 2000.
3. Hsu R.-L., Abdel-Mottaleb M. and Jain A.: *Face detection in color images.* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-706, May 2002.
4. Yang M-H., Kriegman D. and Ahuja N.: *Detecting Faces in Images: A Survey.* IEEE Transactions on Patttern Analysis and Machine Intelligence, Vol 24, n 1, 2002. pp. 34-58
5. Soriano M., Huovinen S., Martinkauppi B. and Laaksonen M.: *Skin Detection in Video under Changing Illumination Conditions.* Proc. 15th International Conference on Pattern Recognition, September 3-8, Barcelona, Spain, Vol 1, pp.839-842, 2000.
6. Fleck M. and Forsyth D. and Bregler C.: *Finding Naked People.* Proceedings of European Conference on Computer Vision , Edited by: Buxton, B.; Cipolla, R. Berlin, Germany: Springer-Verlag, 1996. p. 593-602