A Context-aware Approach for Detecting Skin Colored Pixels in Images

Chandra Mani Sharma
Department of IT
Institute of Tech. & Science
Ghaziabad, INDIA

Saurabh Saxena
Department of IT
Institute of Tech. & Science
Ghaziabad, INDIA

ABSTRACT

Detecting the human skin and its analysis has number of important applications. This is a challenging task as. in images, the skin color is quite sensitive to the chrominance and intensity of the pixels. So the techniques with a single model for skin fail to cope up with the variation in skin colors because of ethnicity, age, lighting etc. This paper proposes a novel technique for skin detection in color images. The proposed technique has two steps; (i) first the faces of humans are detected in the color images (ii) then based on the statistics captured from the sampling of the face area, the rest of the skin is detected. For face detection purpose, we train a binary classifier using machine learning approach. After face detection, the sampled pixels are matched to find the other exposed skin areas using an approach based on Gaussian model for skin.

General Terms

Algorithms, Pattern Recognition, Statistical Models

Keywords

Skin Detection, Image Processing, Gaussian Models, Face Recognition, Multimedia.

1. INTRODUCTION

Skin detection deals with finding skin-colored pixels and regions in an image or in a video sequence. It can be used a preprocessing step towards finding human faces and limbs in images. In images and videos, skin color is an indication of the existence of humans. A skin detector generally transforms a given pixel into an appropriate color space and then uses a skin classifier to label the pixel whether it is a skin or a non-skin pixel. Skin detection means detecting image pixels and regions that contain skin-tone color. Most the research in this area have focused on detecting skin pixels and regions based on their color. Very few approaches attempt to also use texture information to classify skin pixels. Detecting skin-colored pixels has proven quite a challenging task for many reasons. The appearance of skin in an image depends on the illumination conditions (illumination geometry and color) where the image was captured. We humans are very good at identifying object colors in a wide range of illuminations, this is called color constancy. Color constancy is a mystery of perception. Therefore, an important challenge in skin detection is to represent the color in a way that is invariant or at least insensitive to changes in illumination. As will be discussed shortly, the choice of the color space affects greatly the performance of any skin detector and its sensitivity to change in illumination conditions. Another challenge comes from the fact that many objects in the real world might have skin-tone colors. For example, wood, leather, skin-colored clothing, hair, sand, etc. This causes any skin detector to have many false detections in the background if the environment is not controlled. Various techniques have been proposed for skin detection. The skin detection can be done with and without the hardware support. Recently in [1], Tan et al., proposed a skin detection approach that is adaptable to different human skin colors and illumination conditions. In this approach, they combined a smoothed 2-D histogram and Gaussian model, for automatic human skin detection in color images. Zhao et al.[2] proposed a scheme which implements the hardware help in human skin detection. In this, a 0.18 *mu*m triple-well CMOS color processing scheme was proposed. In which on-chip sensor processing enables real-time skin detection. In [3], a texture adaptive skin color detection scheme has been proposed for TV applications, and investigates its feasibility of real-time implementation on a DSP and an FPGA platform. Illumination changes affect the accuracy of skin detection task when a static model is used for skin detection. In order to enhance the accuracy in skin color detection using a static model, the color of an estimated illuminant from images captured under various illumination conditions can be converted to a canonical illuminant[4]. Do et al.[4], proposed a scheme to detect the skin color in varying lighting conditions. In their approach, the illuminant color is estimated from the pixels in the sclera region of the eyes, convert it to a canonical color for skin color detection.

2. SKIN DETECTION APPLICATIONS AND EXAMPLES

Detection of skin in images and videos has many potential applications. One of its early applications has been to identify nude pictures on the Internet to filter the contents [5]. Skin detection is also useful for TV news videos to automatically annotate, retrieve and archive the videos [6]. In an another approach for skin detection to sharpen the pixels for TV news application, Zafarifar et al.,[3] use a color and texture adaptive technique. In such applications, it is general that the face and the hands of the anchor person are the largest skin-tone colored region for a given frame. As news programs are shot in indoor controlled environments where background can be designed not to contain the skin colored objects. In such controlled environments, skin-colored pixels detection can be a useful cue to find humans and limbs in images (or video frames). Skin color can be used as a distinguishing feature for detecting human faces in such controlled environments. As color processing is much faster than processing other facial features, it can be used as a preliminary process for other face detection techniques [7]. Skin detection has also been used to locate body limbs, such as hands, as a part of hand segmentation and tracking systems, e.g., [8]. Forsyth and Fleck[5] demonstrated that skin filter can be used as part of the detection process of images with naked or scantily dressed people. Their technique has three steps. First, a skin filter, based on color and texture, was used to select images with large areas of skin-colored pixels. Then, the output is fed into a geometric filter which identifies the skin-colored regions with cylindrical shapes. Those skin-colored cylinders are grouped into possible human limbs and connected groups of limbs. Images containing sufficiently large skin-colored groups of possible limbs are then reported as containing naked people. Zheng et al. [9] presented an adaptive skin detector for detecting naked pictures on the internet. Their technique applies a face detector on the picture first to find the skin color. They argued that as skin color highly depends on illumination and the race of the person, it is more appropriate to get the skin color from the face of the person in the image. Using the skin color and the property of the texture from the detected face region, the rest of skin pixels in the image can be detected.

3. SKIN DETECTION AND DIFFERENT COLOR SPACES

As was highlighted by Forsyth and Fleck [5] the human skin color has a restricted range of hues and is not deeply saturated, since the appearance of skin is formed by a combination of blood (red) and melanin (brown, yellow). Therefore, the human skin color does not fall randomly in a given color space, but clustered at a small area in the color space. The different color-spaces used for tracking can be classified into the following groups:

(i) **RGB Color Space**

(ii) **TV Color Spaces**

(iii) **Perceptual Color Spaces**

(iv) **Colorimetric Color Spaces**

3.1 RGB Color Space and Skin Detection

It encodes colors as an additive combination of three primary colors: red(R), green (G) and blue (B). RGB Color space is often visualized as a 3D cube where R, G and B are the three perpendicular axes. One main advantage of the RGB space is its simplicity. However, it is not perceptually uniform, which means distances in the RGB space do not linearly correspond to human perception. In addition, RGB color space does not separate luminance and chrominance, and the R,G, and B components are highly correlated. The luminance of a given RGB pixel is a linear combination of the R, G, and B values. Therefore, changing the luminance of a given skin patch affects all the R, G, and B components. In other words, the location of a given skin patch in the RGB color cube will change based on the intensity of the illumination under which such patch was imaged.

3.2 TV Color Spaces and Skin Detection

This is a different class of the orthogonal color spaces, used in TV transmission. It includes YUV, YIQ, and YCbCr. YIQ is used in NTSC TV broadcasting One advantage of using these color spaces is that most video media are already encoded using these color spaces. Transforming from RGB into any of these spaces is a straight forward linear transformation [10]. All these color spaces separate the illumination channel (Y) from two orthogonal chrominance channels (UV, IQ, CbCr). Therefore, unlike RGB, the location of the skin color in the chrominance channels will not be affected by changing the intensity of the illumination. In the chrominance channels the skin color is typically located as a compact cluster with an elliptical shape. This can be seen in Figures 2-d,e,f. This facilitates building skin detectors that are invariant to illumination intensity and that use simple classifiers.

3.3 Perceptual Color Spaces and Skin Detection

Perceptual color spaces, such as HSI, HSV/HSB, and HSL(HLS), have also been popular in skin detection. These color spaces separates three components: the hue (H), the saturation (S) and the brightness (I,V or L). Essentially, HSV-type color spaces are deformations of the RGB color cube and they can be mapped from the RGB space via a nonlinear transformation. One of the advantages of these color spaces in skin detection is that they allow users to intuitively specify the boundary of the skin color class in terms of the hue and saturation. As I, V or L give the brightness information, they are often dropped to reduce illumination dependency of skin color. These spaces have been used by Shin et al. [11] and Albiol et al. [12].

3.4 Colorimetric Color Spaces and Skin Detection

Separating the chromaticity from the brightness is also achieved in Colorimetric color spaces, such as CIE-XYZ, CIE-xy, CIE-Lab defined by the International Commission on Illumination. CIE-XYZ color space is one of the first mathematically defined color space (defined in 1920s). It is based on extensive measurements of human visual perception, and serves as a foundation of many other colorimetric spaces.

4. THE PROPOSED METHOD FOR HUMAN SKIN DETECTION

Detecting human skin in images is a challenging task as the skin color varies from person to person and its very difficult to hypothesize a unique color space representing skin. In our this work the only assumption we have is that we assume the skin color of a person’s face is similar to the color of other limbs. In the proposed approach, the human faces are detected. The color sample is taken to calculate the skin statistics of the person from the detected face. A binary classifier, proposed in our previous work [13], is used for face detection.

A Gaussian model is used for skin detection. This model takes into account the chrominance and illumination of sampled pixels of face areas of humans and matches for the similar pixel statistics in the entire image. The complete procedure of human skin detection is portioned into the following steps:

4.1 Training of Face Detector

Previously, we proposed an object classification scheme in [13]. This multiclass class classifier was used to classify, the objects in a video, into various classes like human objects and the vehicles (cars). The technique was based on machine learning, in which the negative and positive samples were used to train the system. We use a variant of the same technique here. As our previous technique is for multiclass object classification and in case of the binary classification, this technique is similar to Adaptive Boosting [14] and the face classifier proposed by Viola et al.[15]. The classifier construction phase further has sub-phases as listed below:

4.1.1 Collection and Cleansing of Training Data Set

For training the face detector, we collect the positive and negative image data sets. Positive image data sets contain the images of human faces and negative image data set contains the images which are everything but not the human faces. To make our face detector widely applicable, we consider all types all variations in our collection. We include, in our positive image dataset, the faces of all possible age groups (child, adult and old persons) , genders (male and female) and emotions(such as neutral, laughing, sad, angry, staring etc.). In machine learning based approaches for classification, the classifier's accuracy greatly depends on the viability of the data set used for training. The number of images in the positive image data set is 2,700 and in negative data set it is 3,500. After collecting these datasets, the images are converted in a uniform resolution of 40x40. All the inconsistencies in the datasets are resolved at this point only.

Fig.1: Samples in Face Detection Training Data Set

4.1.2 Extracting the Training Features

For performing the training the Haar-like features (have similarity with Haar wavelets) are extracted. For fast extraction of features, we use the concept of Integral Image.

The Integral Image $ii(x,y)$ at location (x,y) contains the sum of all pixels above and left of (x,y) and can be computed in single pass over image using the following pair of equations.

$$\chi(x,y) = \omega(x,y-1) + i(x,y) \quad \textbf{(1)}$$

$$ii(x,y) = ii(x-1,y) + \omega(x,y) \quad \textbf{(2)}$$

where, $\chi(x,y)$ is the cumulative row sum and $i(x,y)$ is the original image.

4.1.3 Training of Face Detector

Let θ be the vector of observation weights and λ is the vector length. We consider that there are M number of stages. in classifier. The final classifier is the liner sum of multiple weak classifiers resulting after each stage. Let $\alpha(\sigma)$ denotes such a typical weak classifier. After the multiple weak classifiers gained after each stage, we get the final classifier which is $\phi(\sigma)$. In this process η^m is the score assigned to m[th] classifier in an stage.

1. Initialize the observation weights $\theta_i = \frac{1}{\lambda}$, i=1,2,………, λ .

2. For $m=1$ to M

(a) Fit a classifier $\alpha^{(m)}(\sigma)$ to the training data using weights θ_i

(b) Compute the misclassification error rate as,

(c) Compute

$$\eta^m = \log\frac{1-\mathrm{err}^{(m)}}{\mathrm{err}^m} + \log(K-1)$$

(d) For i=1,2,….., λ

Set

$$\theta_i \leftarrow \theta_i.\exp(\eta^{(m)} \bullet \delta(c_i \neq \alpha^m(\sigma_i)))$$

(e) Re-normalize the value of θ_i

3. Output

$$\phi(\sigma) = \arg\max_k \sum_{m=1}^{M} \eta^m \bullet \delta(\alpha^m(\sigma) = k)$$

Weak classifiers are linearly combined to form a strong classifier.

4. End

In the present case where there are only two classes viz., face and non-face considering which the classification has to be performed. Here K is the number of classes in the classifier. In case of binary classification problem(such as the present one where there are two classes ; face and non-face), the value of K happens to be 2. So the value of expression log(K-1) this case is 0.

So in case of binary classification the term used in the above algorithm is insignificant and same as the Adaptive Boosting.

The outcome of this training phase is a strong classifier, which is a linear summation of various weaker classifiers generated through the various phases of training.

4.2 Skin Detection

In the first step of the entire process, we detect the faces in the image and take into account the chrominance and illumination of the areas detected as faces. If multiple human faces are From non-smooth regions we imply, the regions of eyes, eye-brows, and others where we cannot sample the skin color

edge detector for detecting the edges on face and eliminate the non-smooth regions. Then, the detected edge pixels are further dilated using a dilation operation to get the optimal non-smooth regions. Finally, we obtain a new image, that only consists of face regions. The Gaussian model is a sophisticated model that is capable of describing complex-shaped distributions and is popular for modeling skin-color distributions. The threshold skin-color distribution in the 2-D histogram is modeled through elliptical Gaussian joint probability distribution functions defined as:

$$p(V\,|\,s)=\sum_{i=1}^{l}\omega_i\; g(c\,|\,\psi_i,\sum_i\ ) \tag{3}$$

Where H is the color vector of (I, B_y), $s=\{\omega_i,\mu_i,\sum_i\ \}$ μ_i is the mean vector, and $\sum_i$ is the diagonal covariance matrix, respectively. ω_i refers to the mixing weights, which satisfy the constraint $\sum_{i=1}^{l}\pi_i=1$. Let (I_n, By_n) be the coordinate of pixel n and is positioned on line D. Distance t of (I_n, By_n) and angle θ are calculated as follows:

$$t=\sqrt{(t_X{}^2\ +\ t_y{}^2\ )} \tag{4}$$

$$\theta=\tan^{-1}(\frac{t_y}{t_x}) \tag{5}$$

$$T_x=\sum_x\sin(\theta) \tag{6}$$

$$T_y=\sum_y\sin(\theta) \tag{7}$$

where $\sum_X$ and $\sum_Y$ are the variance of x axis and y axis for Gaussian model. Distance T is calculated as follows:

detected in an image then multiple arrays are used to hold this statistics. Later on resemblance based search is applied to detect the similarities in the entire image. The detected face region may contain the smooth and non-smooth regions.

correctly. That's why we need to eliminate these regions and avoid sampling these regions before sampling. We use canny

$$T=\sqrt{(T_X{}^2+T_y{}^2)} \tag{8}$$

$$D_g(S_t,\psi,\Sigma)=\begin{cases}1 & if\ T>t\\ 0 & else\end{cases} \tag{9}$$

The resultant image is marked with blue color to show presence of skin-colored pixels. The resultant image, D_g, can be thresholded to get a binary image using equation (9).

5. EXPERIMENTAL RESULTS

We have tested the working of proposed skin detection strategy on various images containing a variety of human skins of different people from different ethnic groups. Generally, it has been observed that the persons who work out or are exposed to outdoor environment has a good mismatch of their face skin and the skin color of their rest of the body parts. So, in the present results we have deliberately chosen the images of sport-persons for experimentation.

In Fig.2, the face and skin detection results have been shown on Afro-American ethnic skin of Serena Williams. The face detector is applied first on the original image (see Fig.2(a)). The sampled face pixels are matched on in the entire image and the resultant image, after skin detection, is shown in Fig.2(c).The presence of skin colored pixels is represented by blue color overlay.

Fig.3 lists the face and skin detection on Indian skin. The face and skin detection results have been demonstrated on the image of Indian Badminton player Saina Nehwal. The method accurately identifies the skinned pixels in the image and ignores the non-skin area.

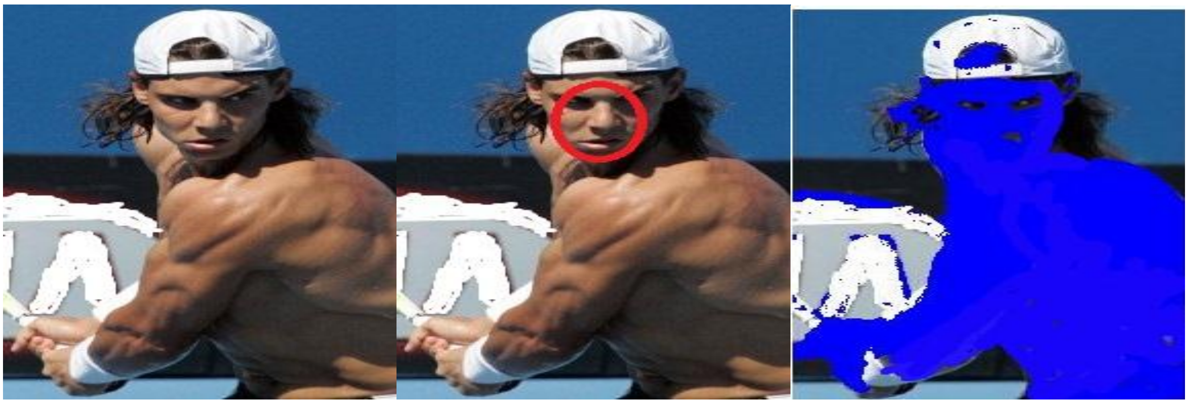
Fig. 4 shows the results with another image. The original image (see Fig. 4(a)) is of Tennis player Rafael Nadal. It is evident from the results that the proposed method accurately performs to detect the skin in the image.

(a) Original Image (b) Face Detection (c) Skin Detection

Fig.2: Face and Skin Detection Results on Afro-American Skin.

(a) Original Image (b) Face Detection (c)Skin Detection

Fig 3: Face and Skin Detection Results on Indian Skin



(a) Original Image (b) Face Detection (c) Skin Detection

Fig.4:Face and Skin Detection on Image an Another Image.

6. FUTURE WORK

Our future work is to devise a mechanism to check the objectionable multimedia contents on web. The social multimedia websites are becoming more commonplace, and posing a threat for minors because of proliferation of objectionable nude images and videos.

Recently, we conducted survey on this theme through a questionnaire comprising of some key questions. The participants of the survey (1000 in number) were from different age groups, genders, and educational qualifications. 55% of the respondents were those who spend more than 10 hours in a week on social networking sites. In this survey we included popular social networking sites Facebook, Orkut, Google+, multimedia content sharing site Youtube, and Google Search. Though, a staggering 64% of participants admitted that these websites are very unsafe for minors, yet 51% of them don't know how to protect them from obscene contents present on these websites. 35% of the participants have no idea about parental control settings, 39% don't know about Google's Safe Search features, 68% don't know about the freeware (such as Pornographic Image Blocker-PickBlock) to keep the Internet Browsing safe for their children. On the question of what they do when encountered with an objectionable content on these sites, 42% say that they ignore it, 33% say they don't mind its presence, only 5% say that they report abuse for this and remaining 10% say that they do nothing. You Tube was rated as the most unsafe to visit by the minors and they are more susceptible to get exposed to the adult contents. While Google+ was rated as the safest website. 53% said that they often encounter with the unexpected objectionable content images while using Image Search on Google and 6% said that they use Safe Search feature of Google to avoid this. 14% of the participants responded that Facebook has the most cumbersome process of reporting the abuse for contents and Orkut has the most convenient mechanism for the same.

7. CONCLUSIONS

The paper proposes an approach for human skin detection. In the proposed technique, face detection is an intermediate step. Detecting and sampling the face skin color enhances the accuracy of the method. The experimental results demonstrate that the proposed method is capable of classifying the skin and non-skin regions in images of persons of different ethnic groups. Skin detection has various useful applications and can be used as a preprocessing step for nudity detection for content filtering on social multimedia sites, which is also our next target. Also, we would like to extend the current work for video objects to have its applicability in wider range of applications.

8. ACKNOWLEDGEMENTS

We are thankful to all the respondents of our online quiz survey on-"ARE SOCIAL MULTIMEDIA SITES SAFE FOR MINORS?" .

9. REFERENCES

[1] Wei Ren Tan; Chee Seng Chan; Yogarajah, P.; Condell, J., "A Fusion Approach for Efficient Human Skin Detection," IEEE Trans. On Industrial Informatics, Vol. 8, Issue 1, pp. 138-147, 2012.

[2] Xiaojin Zhao; Boussaid, F.; Bermak, A.," Characterization of a 0.18 µ m CMOS Color Processing Scheme for Skin Detection," IEEE Sensors Journal, Vol. 7, Issue 11, pp. 1471-1474, 2007.

[3] Zafarifar, B.; van den Kerkhof, T.; de With, P.H.N. "Texture-adaptive skin detection for TV and its real-time implementation on DSP and FPGA," IEEE Trans. On Consumer Electronics, Vol. 58, Issue 1, pp. 161-169, 2012.

[4] Hyun-Chul Do; Ju-Yeon You; Sung-Il Chien, "Skin Color Detection through Estimation and Conversion of Illuminant Color Under Various Illumination," IEEE Trans. On Consumer Electronics, Vol. 53, Issue 3, pp. 1103-1108, 2007.

[5] Fleck, M.M., Forsyth, D.A., Bregler, C.: Finding naked people. In: Proceedings of the European Conference on Computer Vision (ECCV). Pp. 593-602, 1996.

[6] Abdel-Mottaleb, M., Elgammal, A.: Face detection in complex environments from color\ lmages. In: Proceedings of the International Conference on Image Processing (ICIP), pp. 622-626, 1996.

[7] Senior, A., Hsu, R.L., Mottaleb, M.A., Jain, A.K.: Face detection in color images. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 24, No.5, pp. 696-706, 2002.

[8] Imagawa, K., Lu, S., Igi, S.: Color-based hands tracking system for sign language recognition. In: FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition, Washington, DC, USA, IEEE Computer Society (1998) 462

[9] Zheng, Q.F., Zhang, M.J., Wang, W.Q.: A hybrid approach to detect adult web images. In: PCM (2), pp. 609-616, 2004.

[10] Burger, W., Burge, M.: Digital Image Processing, an Algorithmic Introduction Using Java. Springer (2008)

[11] Shin, M.C., Chang, K.I., Tsap, L.V.: Does colorspace transformation make any difference on skin detection? In: WACV '02: Proceedings of the Sixth IEEEWorkshop on Applications of Computer Vision, Washington, DC, USA, IEEE Computer Society (2002) 275

[12] Albiol, A., Torres, L., Delp, E.: Optimum color spaces for skin detection. In: Proceedings of the International Conference on Image Processing (ICIP), pp. 122-124, 2001.

[13] C.M. Sharma, A.K.S. Kushwaha, R. Roshan, R. Porwal, A. Khare,"Intelligent Object Classification Scheme using Machine Learning and Offline Feature Based Technique," Int. J. Computer Science Issues, Vol.9, Issue 1, No.3, pp.,247-256, 2012.

[14] J. Zhu, H. Zou, S. Rosset, and T. Hastie, " Multiclass adaboost, " Int. J. of Statistics and Its Interface, vol. 2, pp. 349-360, 2009.

[15] P. Viola and M. Jones, " Robust real-time object detection," Int. J. of Computer Vision, vol. 57, no. 2, pp. 137–154, 2001.