Skin Detection - a Short Tutorial †

Ahmed Elgammal, Crystal Muang and Dunxu Hu
Department of Computer Science, Rutgers University, Piscataway, NJ, 08902, USA

Skin detection is the process of finding skin-colored pixels and regions in an image or a video. This process is typically used as a preprocessing step to find regions that potentially have human faces and limbs in images. Several computer vision approaches have been developed for skin detection. A skin detector typically transforms a given pixel into an appropriate color space and then use a skin classifier to label the pixel whether it is a skin or a non-skin pixel. A skin classifier defines a decision boundary of the skin color class in the color space based on a training database of skin-colored pixels.

Introduction

Skin color and textures are important cues that people use consciously or unconsciously to infer variety of culture-related aspects about each other. Skin color and texture can be an indication of race, health, age, wealth, beauty, etc. [1]. However, such interpretations vary across cultures and across the history. In images and videos, skin color is an indication of the existence of humans in such media. Therefore, in the last two decades extensive research have focused on skin detection in images. 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.

As will be described shortly, detecting skin pixels are rather computationally easy task and can be done very efficiently, a feature that encourages the use of skin detection in many video analysis applications. For example, in one of the early applications, detecting skin-colored regions was used to identify nude pictures on the internet for the sake of content filtering [2]. In another early application, skin detection was used to detect anchors in TV news videos for the sake of video automatic annotation, archival, and retrieval [3]. In such an application, it is typical that the face and the hands of the anchor person are the largest skin-tone colored region in a given frame since, typically, news programs are shot in indoor controlled environments with man-made background materials that hardly contain skin-colored objects. In many similar applications, where the background is controlled or unlikely to contain skin-colored regions, detecting skin-colored pixels can be a very efficient cue to find human faces and hands in images. An example in the context of biometric is detecting faces for face recognition in an controlled environment.

†To appear as an entry in Encyclopedia of Biometrics by Springer-Verlag Berlin Heidelberg 2009

Detecting skin-colored pixels, although seems a straightforward easy task, has proven quite challenging 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.

A Framework for Skin Detection

Skin detection process has two phases: a training phase and a detection phase. Training a skin detector involves three basic steps:

1. Collecting a database of skin patches from different images. Such a database typically contains skin-colored patches from a variety of people under different illumination conditions.

2. Choosing a suitable color space.

3. Learning the parameters of a skin classifier.

Given a trained skin detector, identifying skin pixels in a given image or video frame involves:

1. Converting the image into the same color space that was used in the training phase.

2. Classifying each pixel using the skin classifier to either a skin or non-skin.

3. Typically post processing is needed using morphology to impose spatial homogeneity on the detected regions.

In any given color space, skin color occupies a part of such a space, which might be a compact or large region in the space. Such region is usually called the skin color cluster. A skin classifier is a one-class or two-class classification problem. A given pixel is classified and labeled whether it is a skin or a non-skin given a model of the skin color cluster in a given color space. In the context of skin classification, true positives are skin pixels that the classifier correctly labels as skin. True negatives are non-skin pixels that the classifier correctly labels as non-skin. Any classifier makes errors: it can wrongly label a non-skin pixel as skin or a skin pixel as a non-skin. The former type of errors is referred to as false positives (false detections) while the later is false negatives. A good classifier should have low false positive and false negative rates. As in any classification problem, there is a tradeoff between false positives and false negatives. The more loose the class boundary, the less the false negatives and the more the false positives. The tighter the class boundary, the more the false negatives and the less the false positives. The same applies to skin detection. This

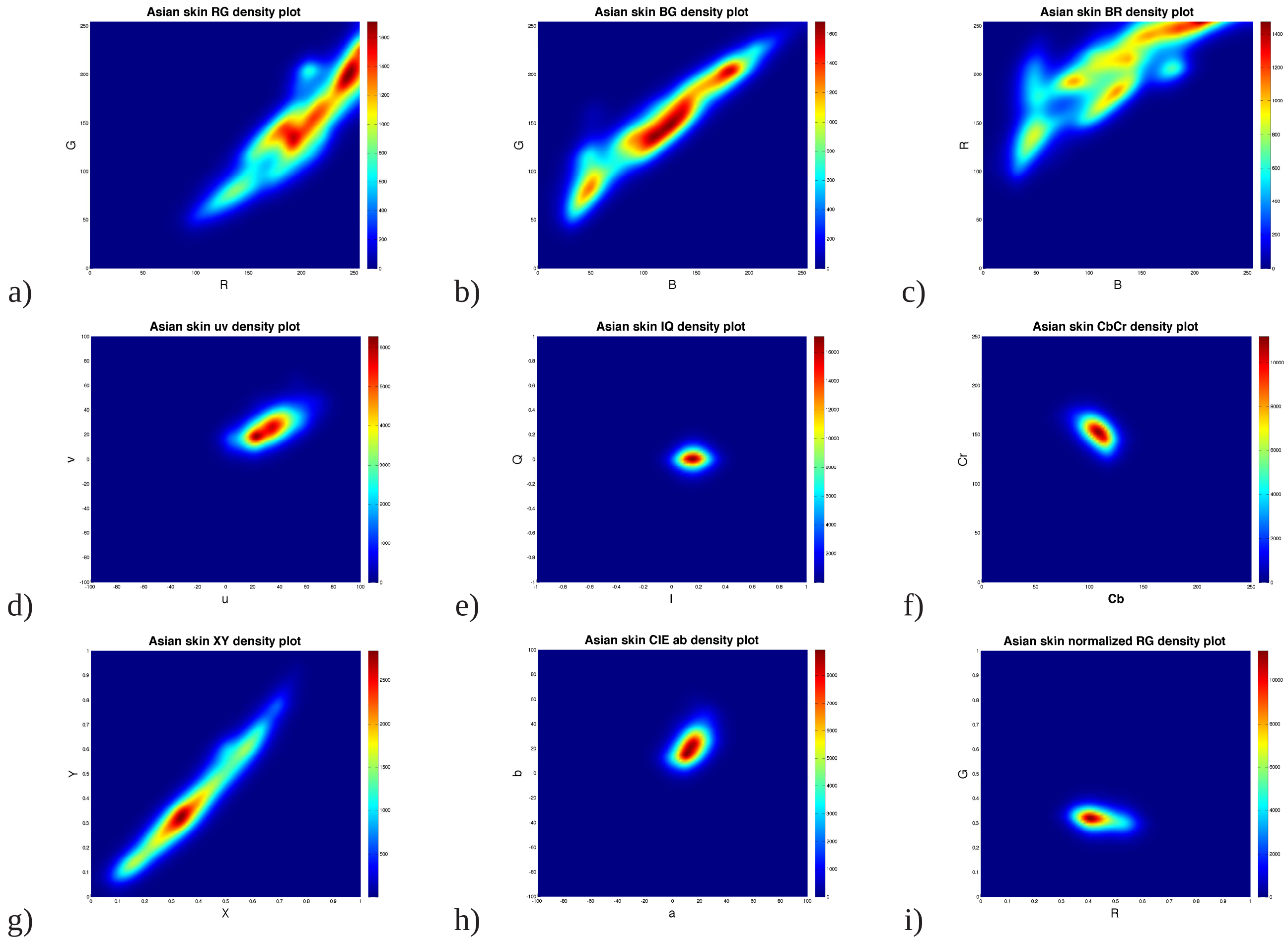

Fig. 2: Density plots of Asian skin in different color spaces

makes the choice of the color space extremely important in skin detection. The color needs to be represented in a color space where the skin class is most compact in order to be able to tightly model the skin class. The choice of the color space directly affects the kind of classifier that should be used.

Skin Detection and Color Spaces

As was highlighted by Forsyth and Fleck [2] 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. But it is not the same for all the color spaces. Variety of color spaces have been used in skin detection literature with the aim of finding a color space where the skin color is invariant to illumination conditions. The choice of the color

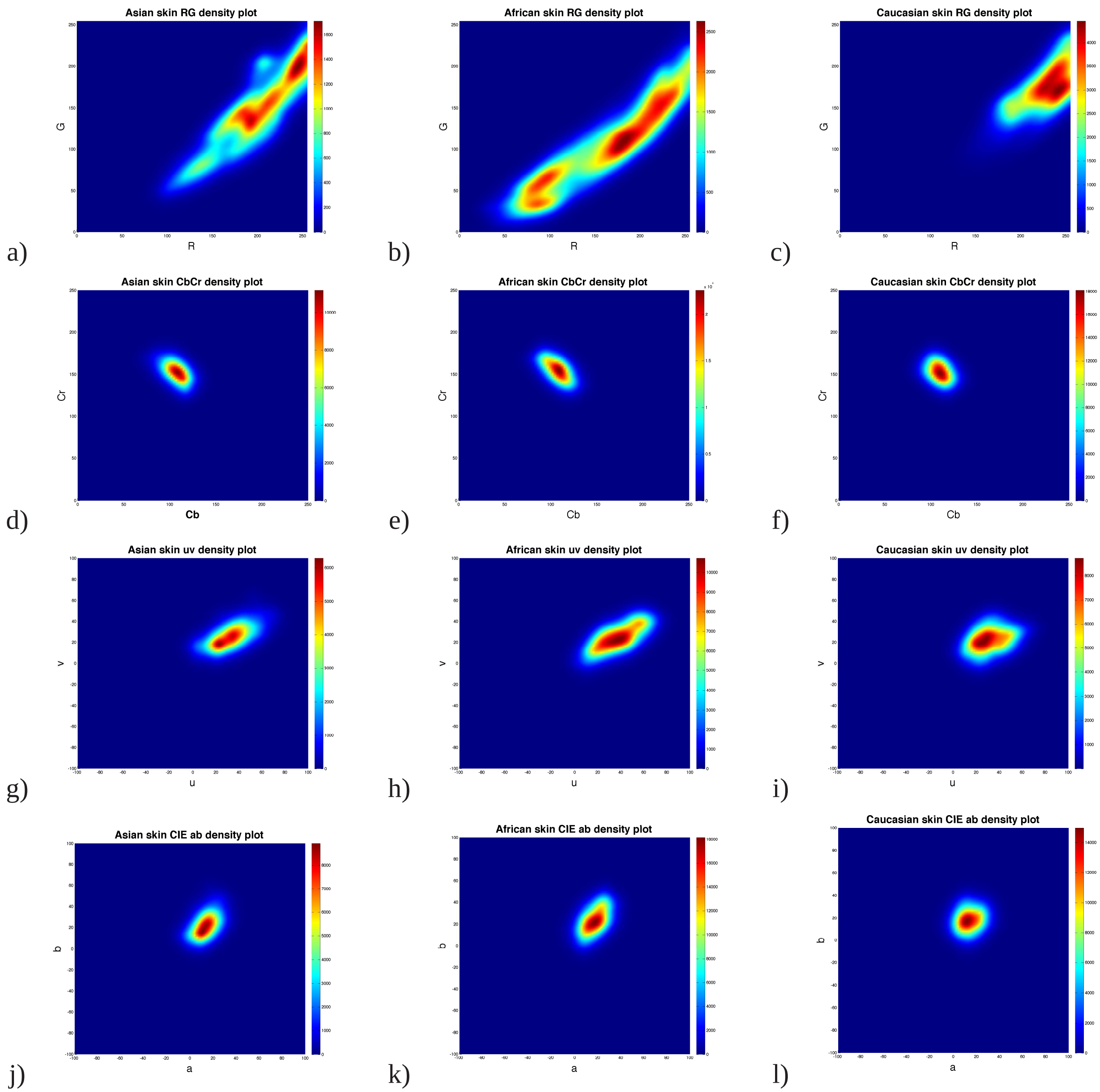

Fig. 3: Density plots of Asian, African and Caucasian skin in different color spaces

spaces affects the shape of the skin class, which affects the detection process. Here, some color spaces, which are typically used in skin detection, are briefly described, and the way they affect the skin detection is discussed. The goal of the discussion it to highlight answers to the following questions:

- Given a skin patch, where will it be located in a given color space?
- Given a skin patch, what effect will changing the illumination intensity have in its location in a given color space?
- Given skin patches from different people from the same race, how are all these patches related in a given color space?
- Given skin patches from different people races, how are all these patches related in a given color space?

Figures 2 and 3 help illustrate the answers for these questions. Figure 2 shows density plots for skin-colored pixels obtained from images of different Asian people plotted in different color spaces. Figure 3 shows density plots for skin-colored pixels from different people from different races: Asian, African and Caucasian plotted in different color spaces.

RGB Color Space and Skin Detection:

$\mathbf{RGB}$ color space is the most commonly used color space in digital images. It encodes colors as an additive combination of three primary colors: red(R), green (G) and blue (B). $\mathbf{RGB}$ Color space is often visualized as a 3D cube where R, G and B are the three perpendicular axes. One main advantage of the $\mathbf{RGB}$ space is its simplicity. However, it is not perceptually uniform, which means distances in the $\mathbf{RGB}$ space do not linearly correspond to human perception. In addition, $\mathbf{RGB}$ color space does not separate luminance and chrominance, and the R,G, and B components are highly correlated. The luminance of a given $\mathbf{RGB}$ 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 $\mathbf{RGB}$ color cube will change based on the intensity of the illumination under which such patch was imaged! This results in a very stretched skin color cluster in the $\mathbf{RGB}$ color cube. This can be noticed in the first row of Figure 2 where skin patches from images of Asian people taken at arbitrary random illumination are plotted in the $\mathbf{RGB}$ space. The skin color cluster is extended in the space to reflect the difference illumination intensities in the patches. Similarly, the skin color clusters for patches from different races will be located at different locations in the $\mathbf{RGB}$ color space. This can be seen in the first row of Figure 3. Despite these fundamental limitations, $\mathbf{RGB}$ is extensively used in skin detection literature because of its simplicity. For example, $\mathbf{RGB}$ is used by Rehg and Jones [4] and yield quite satisfying performance.

TV Color Spaces and Skin Detection:

A different class of color spaces are the orthogonal color spaces used in TV tansmission. This includes $\mathbf{YUV}$, $\mathbf{YIQ}$, and $\mathbf{YC}_b\mathbf{C}_r$. $\mathbf{YIQ}$ is used in NTSC TV broadcasting while $\mathbf{YC}_b\mathbf{C}_r$ is used in JPEG image compression and MPEG video compression. One advantage of using these color spaces is that most video media are already encoded using these color spaces. Transforming from $\mathbf{RGB}$ into any of these spaces is a straight forward linear transformation [5]. All these color spaces separate the illumination channel ($\mathbf{Y}$) from two orthogonal chrominance channels ($\mathbf{UV}$, $\mathbf{IQ}$, $\mathbf{C}_b\mathbf{C}_r$). Therefore, unlike $\mathbf{RGB}$, 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. The density of the skin color over the chrominance channels can be easily approximated using a multivariate Gaussian distribution. Moreover, the skin colors of different races almost co-locate in the chrominance channels. This can be seen in the second and third rows of Figures 3. Therefore, using such color spaces results in skin detectors which are invariant to human race. The simplicity of the transformation and the invariant properties made such spaces widely used in skin detection applications [3, 6, 7, 8, 9, 10].

Perceptual Color Spaces and Skin Detection:

Perceptual color spaces, such as $\mathbf{HSI}$, $\mathbf{HSV}/\mathbf{HSB}$, and $\mathbf{HSL}$($\mathbf{HLS}$), 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, $\mathbf{HSV}$-type color spaces are deformations of the $\mathbf{RGB}$ color cube and they can be mapped from the $\mathbf{RGB}$ 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 $\mathbf{I}$, $\mathbf{V}$ or $\mathbf{L}$ give the brightness information, they are often dropped to reduce illumination dependency of skin color. These spaces have been used by Shin et al. [6] and Albiol et al. [8].

Colorimetric Color Spaces and Skin Detection:

Separating the chromaticity from the brightness is also achieved in Colorimetric color spaces, such as $\mathbf{CIE}$-$\mathbf{XYZ}$, $\mathbf{CIE}$-xy, $\mathbf{CIE}$-Lab defined by the International Commission on Illumination (Commission Internationale d'Èclairage - CIE). $\mathbf{CIE}$-$\mathbf{XYZ}$ 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. $\mathbf{CIE}$-$\mathbf{XYZ}$ can be achieved through a linear coordinate transformation of the $\mathbf{RGB}$ color space. The Y component corresponds to the lightness of the color (the luminance). The chromaticity values (x, y) can be achieved by central projection into the plane $X + Y + Z = 1$ and then projecting into the XY plane. For details see [5]. The result is the well-known horse-shaped $\mathbf{CIE}$-xy chromaticity diagram defining the hue and saturation of any color. One of the disadvantages of the $\mathbf{XYZ}$ and the xy color spaces is that the color differences are not perceived equally in different regions of the color space. In contrast, the $\mathbf{CIE}$-Lab separates a luminance variable L from two perceptually uniform chromaticity variables (a,b) Figure 2-h shows the skin color density for Asian skin in the a,b chromaticity space. Figure 3 (last row) shows the skin color density for different races in the a,b space. Despite the many advantages of such color spaces, they are rarely used in skin detection. This is mainly because the transformation from $\mathbf{RGB}$ is more computationally expensive than other spaces. $\mathbf{CIE\text{-}XYZ}$ color space was used by Shin et al. [6] in comparison with other color spaces. The chrominance xy plane was used by Lee and Yoo [9].

Skin Classifiers

A variety of classification techniques have been used in the literature for the task of skin classification. A skin classifier is a one-class classifier that defines a decision boundary of the skin color class in a feature space. The feature space in the context of skin detection is simply the color space chosen. Any pixel which color falls inside the skin color class boundary is labeled as skin. Therefore, the choice of the skin classifier is directly induced by the shape of the skin class in the color space chosen by a skin detector. The more compact and regularly shaped the skin color class, the more simple the classifier.

The simplest way to decide whether a pixel is skin color or not is to explicitly define a boundary. Brand and Mason [11] constructed a simple one-dimensional skin classifier: a pixel is labeled as a skin if the ratio between its R and G channels is between a lower and an upper bound. They also experimented with one-dimensional threshold on IQ plane of YIQ space where the "I" value is used for thresholding. Other methods explicitly defines the skin color class boundary in a two-dimensional color space using elliptical boundary models [9]. The parameters of the elliptical boundary can be estimated from the skin database at the raining phase.

Baysian Approach for Skin Detection: Skin classification can be defined probabilistically as: given a pixel with color c what is the probability of it being skin pixel $P(skin|c)$. Once this probability is computed, the pixel is labeled as a skin pixel if such probability is larger than a threshold and non-skin otherwise. Obviously we can not compute such probabilities for every possible color (e.g., in 24 bit $\mathbf{RGB}$, there are 256^3 colors). Fortunately, using Bayes rule, this can be rewritten as

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|notskin)P(notskin)}$$

Bayes rule defines the posterior probability of a pixel being skin given its color ($P(skin|c)$) in terms of the likelihood of observing such color given the skin class ($P(c|skin)$) and the prior probability of the skin class $P(skin)$. The prior probability measures our guess about a random pixel being a skin without observing its color. The denominator in the Bayes rule is the total probability of observing the color c, a factor that does not affect the decision whether a pixel ought to be labeled as skin or non-skin. Given Bayes rule, the skin classification reduces to computing the likelihood term, i.e., $P(c|skin)$. Given a database of skin-colored pixels we can estimate the probability density function (pdf) of $P(c|skin)$. Several approaches have been introduced to compute this pdf including the use of histograms [4], the use of a single Gaussian model, or a Mixture of Gaussians model [12] to approximate such pdf.

The skin classifier can also be posed as a two-class problem. From Bayes rule, this results in computing the likelihood ratio of observing a given color given a skin class versus a nonskin class, i.e., $P(c|skin)/P(c|notskin)$. Such ratio can then be thresholded to decide whether a pixel is a skin or non-skin pixel. Besides modeling the likelihood of an observed color given the skin class, the complementary class needs to be models. That is, modeling the probability density function of non-skin pixels $P(c|notskin)$. Rehg and Jones [4] approximated such pdfs using 3D histograms in the $\mathbf{RGB}$ space based on a large database of skin and non-skin images

Skin Detection Applications and Examples

Human face localization and detection is the first step in obtaining face biometrics. Skin color is a distinguishing feature of human faces. In a controlled background environment, skin detection can be sufficient to locate faces in images. As color processing is much faster than processing other facial features, it can be used as a preliminary process for other face detection techniques [10]. Skin detection has also been used to locate body limbs, such as hands, as a part of hand segmentation and tracking systems, e.g., [13].

Forsyth and Fleck[2] 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. [7] 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.

Skin Detection Performance

Regardless of the choice of the color space and the classification method, most published research on skin detection reports about 95% true detection while the false detection rates varies from 15%-30%.

Summary

Skin detection in color images and videos is a very efficient way to locate skin-colored pixels, which might indicate the existence of human faces and hands. However, many objects in the real world have skin-tone colors, such as some kinds of leather, sand, wood, fur, etc., which might be mistakenly detected by a skin detector. Therefore, skin detection can be very useful in finding human faces and hands in controlled environments where the background is guaranteed not to contain skin-tone colors. Since skin detection depends on locating skin-colored pixels, its use is limited to color images, i.e., it is not useful with gray-scale, infrared, or other types of image modalities that do not contain color information. There have been extensive research on finding human faces in images and videos using other cues such as finding local facial features or finding holistic facial templates [14]. Skin detection can also be used as an efficient preprocessing filter to find potential skin regions in color images prior to applying more computationally expensive face or hand detectors.

References

[1] Bernhard Fink, K.G., Matts, P.J.: Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. Evolution and Human Behavior **27**(6) (2006) 433–442

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

[3] Abdel-Mottaleb, M., Elgammal, A.: Face detection in complex environments from color lmages. In: Proceedings of the International Conference on Image Processing (ICIP). (1999) 622–626

[4] Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International Journal of Computer Vision (IJCV) **46**(1) (2002) 81–96

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

[6] 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 IEEE Workshop on Applications of Computer Vision, Washington, DC, USA, IEEE Computer Society (2002) 275

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

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

[9] Lee, Y., Yoo, S.I.: An elliptical boundary model for skin color detection. In: Proc.of the Int. Conf. on Imaging Science, Systems, and Technology. (2002)

[10] 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) **24**(5) (2002) 696–706

[11] Brand, J., Mason, J.: A comparative assessment of three approaches to pixellevel human skin-detection. **1** (2000) 1056–1059

[12] Yang, M., Ahuja, N.: Gaussian mixture model for human skin color and its application in image and video databases. In: In Proc. of the SPIE: Conference on Storage and Retrieval for Image and Video Databases (SPIE 99). Volume 3656. (1999) 458466

[13] 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

[14] Yang, M., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **24**(1) (2002) 34–58