REGION-BASED SKIN COLOR DETECTION

Rudra P K Poudel[1], Hammadi Nait-Charif[1], Jian J Zhang[1] and David Liu[2]
[1]*Media School, Bournemouth University, Weymouth House, Talbot Campus, BH12 5BB, UK*
[2]*Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA*
{*rpoudel, jzhang, hncharif*}*@bournemouth.ac.uk, David-Liu@siemens.com*

Keywords: skin-color detection, region-based, superpixels, Bayes classifier, conditional random field

Abstract: Skin color provides a powerful cue for complex computer vision applications. Although skin color detection has been an active research area for decades, the mainstream technology is based on the individual pixels. This paper presents a new region-based technique for skin color detection which outperforms the current state-of-the-art *pixel-based* skin color detection method on the popular *Compaq dataset* (Jones and Rehg, 2002). Color and spatial distance based clustering technique is used to extract the regions from the images, also known as *superpixels*. In the first step, our technique uses the state-of-the-art non-parametric pixel-based skin color classifier (Jones and Rehg, 2002) which we call the basic skin color classifier. The pixel-based skin color evidence is then aggregated to classify the superpixels. Finally, the *Conditional Random Field* (CRF) is applied to further improve the results. As CRF operates over superpixels, the computational overhead is minimal. Our technique achieves 91.17% true positive rate with 13.12% false negative rate on the Compaq dataset tested over approximately 14,000 web images.

1 INTRODUCTION

Skin color provides a powerful cue in complex computer vision applications such as hand tracking, face tracking, and pornography detection. Skin color detection is computationally efficient yet invariant of rotation, scaling and occlusion. These are the major reasons for its popularity. The main challenges of skin color detection are illumination, ethnicity background, make-up, hairstyle, eyeglasses, background color, shadows and motion (Kakumanu et al., 2007). Many of the skin color detection problems could be overcome by using *infrared* (Socolinsky et al., 2003) and *spectral imaging* (Pan et al., 2003). However, such systems are expensive as well as cumbersome to implement. Moreover, there are many situations where such systems can not be used such as image retrieval on the internet.

Most of the skin color detection methods are *pixel-based*, which treat each skin or non-skin pixel individually without considering its neighbours. However, it is natural to treat skin or non-skin as regions instead of individual pixels. Surprisingly, there are only few region-based skin detection techniques (Yang and Ahuja, 1998), (Kruppa et al., 2002), (Jedynak et al., 2003) and (Sebe et al., 2004). Kruppa (Kruppa et al., 2002), Yang and Ahuja (Yang and Ahuja, 1998) searched for elliptical skin color shape to find the face. Sebe (Sebe et al., 2004) used fixed 3x3 pixel patches to train a Bayesian network and Jedyank (Jedynak et al., 2003) smoothed the results using hidden Markov model. This paper proposes a new technique purely based on the concept of regions, irrespective of the underlying geometrical shape. As such, this technique can be easily integrated into any skin detection based system.

Our technique uses a segmentation technique called *superpixel* (Moore et al., 2008) and (Ren and Malik, 2003) to group similar color pixels together. Then each superpixel is classified as skin or non-skin by aggregating pixel-based evidence obtained using a histogram based Bayesian classifier similar to (Jones and Rehg, 2002). The result is further improved with CRF, which operate over superpixels instead of pixels. Even though the segmentation cost is an overhead over the pixel-based approach, it greatly reduces the processing cost further down the line, such as smoothing with CRF. Also, aggregation of pixels into regions helps to reduce local redundancy and the probability of merging unrelated pixels (Soatto, 2009). Since superpixels preserve the boundary of the objects, it helps to achieve very accurate object segmentations

(Fulkerson et al., 2009).

The presented method not only outperforms the current state-of-the-art pixel-based skin color detection methods but also extracts larger skin regions while still keeping the false-positive rate lower, providing semantically more meaningful results. This could in turn benefit higher-level vision tasks, such as face or hand detection. Related work is discussed in section 2; section 3 presents the proposed region-based skin color detection technique; experiments and results are discussed in section 4. Finally, we summarize our work in section 5.

2 RELATED WORK

Skin color detection has two important parts: one is color space selection and another is color modelling. RGB: (Jones and Rehg, 2002), HSV: (Zhu et al., 2004), CIE-Lab: (Kawato and Ohya, 2002), YCbCr: (Wong et al., 2003), and normalized RGB: (Brown et al., 2001) are popular color spaces, with RGB and HSV being the most frequently used. CIE-Lab uniformly represents the color based on how two colors differ to the human observer. HSV shows better results under varying illumination (Kakumanu et al., 2007). However, most camera output RGB and illumination variation can be eliminated by increasing sample size (Jones and Rehg, 2002). Due to this reason the RGB color space is chosen in our experiments.

Skin color modelling falls in three categories: explicitly defined skin region (Peer et al., 2003), non-parametric and parametric methods. Histogram based Bayes classifier is a popular non-parametric modelling approach. Jones and Rehg (Jones and Rehg, 2002) used RGB color space and histograms based Bayes classifier and obtained 90% true positive rate with 14.5% false positive rate on unconstrained web images, a dataset made up of approximately 14,000 images. On parametric skin modelling technique, mixture of Gaussian has shown the best result (Terrillon et al., 2000). However, Jones and Rehg (Jones and Rehg, 2002) showed that given enough samples, the histogram based Bayes classifier technique is slightly better than mixture of Gaussian. Neural Network (Phung et al., 2002), self organizing map (Brown et al., 2001), Bayesian network (Sebe et al., 2004) and a few other methods have been used for skin color modelling.

This paper presents a region-based skin color detection method with no prior knowledge on the geometric shape of the regions. The works of Yang and Ahuja (Yang and Ahuja, 1998), Kruppa (Kruppa et al., 2002), Jedyank (Jedynak et al., 2003) and Sebe (Sebe et al., 2004) are the close to ours. However, Yang and Ahuja (Yang and Ahuja, 1998) used multi-scale segmentations to find elliptical regions for face detection. Hence, their model is biased toward elliptical objects. Kruppa (Kruppa et al., 2002) also used a similar concept to find the elliptical region using color and shape information for face detection. Sebe (Sebe et al., 2004) used 3x3 fixed size pixel patches. Our presented technique uses patches with varying sizes, which is purely based on image evidence, i.e. skin color in this case. Also, Jedyank (Jedynak et al., 2003) used hidden Markov model at pixel level, while we use conditional random fields and operate on superpixel, as described in section 3.4.

(a) Original image (b) After segmentation (c) Segmentation visualization

Figure 1: An example of superpixel segmentation.

3 REGION-BASED APPROACH

We argue that skin is better presented as regions rather than individual pixels. The proposed region-based approach has four major components: *basic skin classifier* (section 3.1), extraction of regions called superpixels (section 3.2), superpixels classification (section 3.3), and a smoothing procedure with CRF (section 3.4). Each step is discussed in detail below.

3.1 Basic Skin Color Classifier

Any good skin color classification method can be used as a basic skin color classifier. This paper uses the histogram based Bayesian classifier similar to that of Jones and Rehg (Jones and Rehg, 2002), a state-of-the-art skin color detection technique.

Learning skin and non-skin histograms: densities of skin and non-skin color *histograms* are learned from the *Compaq dataset* (Jones and Rehg, 2002). The Compaq skin color dataset has approximately 4,700 skin images and 9,000 non-skin images collected from free web crawling. It has images from all ethnic groups with uncontrolled illumination and background conditions. The number of manually labelled pixels is nearly 1 billion. Skin and non-skin histograms are obtained in RGB color space with 32

bins for each color channel, exactly same to the settings in Jones and Rehg (Jones and Rehg, 2002). Equal number of skin images are randomly divided for training and testing. Similarly, equal number of non-skin images are randomly divided for training and testing.

Bayesian skin classifier: Naive Bayes is used to build the skin and non-skin classifier. The probability of a color being skin s given a color c, $P(s|c)$, is given by

$$P(s|c) = \frac{P(c|s)P(s)}{P(c)} \quad (1)$$

where, $P(c|s)$ is the likelihood of a given color c being skin, $P(s)$ is skin color prior and $P(c)$ is color prior. Similarly, the probability of a color being non-skin ns given a color c is given by

$$P(ns|c) = \frac{P(c|ns)P(ns)}{P(c)} \quad (2)$$

where, $P(c|ns)$ is the likelihood of a given color c being non-skin and $P(ns)$ prior for non-skin color. Further $P(c)$ could be calculated as following

$$P(c) = P(c|s)P(s) + P(c|ns)P(ns) \quad (3)$$

$P(c|s)$ and $P(c|ns)$ are directly calculated from skin and non-skin histograms. Prior probabilities: $P(s)$ and $P(ns)$ can also estimated from the total number of skin and non-skin samples in the training dataset. However, for skin and non-skin classification, we can simply compare $P(s|c)$ to $P(ns|c)$. Using equations (1) and (2), the ratio of $P(s|c)$ to $P(ns|c)$ can be simplified to

$$\frac{P(s|c)}{P(ns|c)} = \frac{P(c|s)P(s)}{P(c|ns)P(ns)} \quad (4)$$

Equation (4) can be thresholded to produce a skin and non-skin classification rule. Further, $P(s)$ and $P(ns)$ are also constant so this can be simplified as follows

$$\frac{P(c|s)}{P(c|ns)} > \Theta \quad (5)$$

where Θ is a constant threshold value.

In the experiments, equation (5) is used to find the skin and non-skin probability for pixels. The values of $P(c|s)$ and $P(c|ns)$ are directly looked-up from normalized skin and non-skin histograms respectively.

3.2 Superpixels

A region or collection of pixels is called a superpixel. A five dimensional vector is used to extract the superpixels: three RGB color channels and two positional coordinates of the pixel, using the *quick shift* (Vedaldi and Soatto, 2008) image segmentation algorithm. Superpixels generated from this approach vary in size and shape, hence the number of superpixels in each image is highly dependent upon the complexity of the image. An image with low color variation will have a smaller number of superpixels than an image with high color variation, as there is no penalty for boundary violation. Generally, the concept of boundary is not used when extracting the superpixels, however different objects have different texture or color which will implicitly act as boundaries. Figure 1 shows the example of superpixels of an image. In our work we have used the superpixel extraction library (Vedaldi and Fulkerson, 2008) for superpixel segmentation.

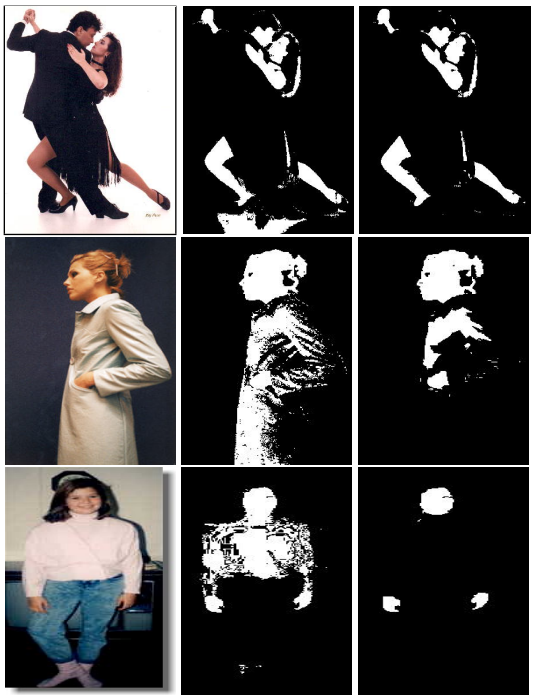


Figure 2: Comparison between pixel-based (Jones and Rehg, 2002) (middle) and region-based with CRF(right) skin color classification techniques.

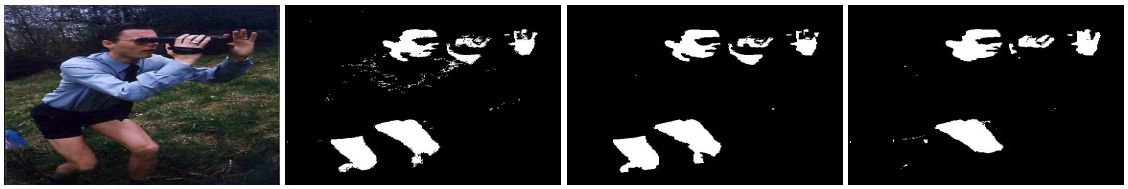
3.3 Superpixel Classification

First, the pixel based skin color classifier defined on section 3.1 is used to classify the pixels of the images. Then the probability of being skin for a given superpixel sp with N number of color pixels c is defined as follows

$$P(s|sp) = \frac{1}{N}\sum_{i}^{N} P(s|c_i) \quad (6)$$

Similarly the probability of being non-skin for a given superpixel sp with N number of color pixels c is defined as follows

$$P(ns|sp) = \frac{1}{N}\sum_{i}^{N} P(ns|c_i) \quad (7)$$



(a) Original image (b) Pixel-based (c) Region-based without CRF (d) Region-based with CRF

Figure 3: This example shows the advantages of the region-based approach even without CRF (see sub figures b and c). Sub figures c and d show the failure case when CRF is applied.

3.4 Smoothing with CRF

Skin regions have varying size and shape, depending upon the camera angle, distance from the camera and human body factors. Hence, to obtain smooth skin regions but still preserve the skin and non-skin boundaries, it is necessary to introduce some constraints. CRF provides a natural way of combining pairwise constraints. Color difference and length of boundary between adjacent superpixels are used as pairwise constraints. Optimum skin and non-skin labelling L of all superpixels S of an image is defined as follows

$$-\log(P(L|S;\omega)) = -\sum_{s_i \in S} \Psi(l_i|s_i) + \omega \sum_{(s_i,s_j)\in E} \Phi(c_i,c_j|s_i,s_j) \quad (8)$$

where ω is the weight of pairwise constraint, E is the set of edges of superpixel, and i and j are index nodes in superpixel level graph of an image.

Color potential ($\Psi(l_i|s_i)$): the color potential Ψ captures the skin and non-skin probability of superpixel s_i. We have used skin and non-skin probability for superpixel directly from superpixel classification defined in section 3.3 for color potential Ψ as follows

$$\Psi(l_i|s_i) = \log(P(l_i|s_i)) \quad (9)$$

Edge and boundary potential ($\Phi(c_i,c_j|s_i,s_j)$): pairwise edge and boundary potential Φ is defined similar to those of (Fulkerson et al., 2009)

$$\Phi(c_i,c_j|s_i,s_j) = \left(\frac{L(s_i,s_j)}{1+||s_i-s_j||}\right), [c_i \neq c_j] \quad (10)$$

where $L(s_i,s_j)$ is the shared boundary length, and $||s_i-s_j||$ is the Euclidean norm of the color difference between s_i and s_j superpixels.

Only one pairwise potential is used to make the system as simple as possible to show that treating skin color with regions is more effective than with pixels. This implementation has only one weighting factor ω, which is optimized using cross validation. We use the multi-label graph optimization library of (Boykov et al., 2001), (Boykov and Kolmogorov, 2004) and (Kolmogorov and Zabih, 2004) for the inference of skin and non-skin regions. CRF graph is built on the superpixel level hence CRF optimization is fast.

Method	TP	FP
Jones and Rehg (2002)	90%	14.2%
Our (superpixel only)	**91.44%**	**13.73%**
Our (superpixel and CRF)	**91.17%**	**13.12%**

Table 1: The results of pixel-based and our region-based technique.

4 EXPERIMENTS AND RESULTS

Equal number of training and testing sets are randomly chosen from the Compaq dataset (Jones and Rehg, 2002) and same training and testing sets are used for all experiments. The Compaq dataset has approximately 4,700 skin and 9,000 non-skin images, freely collected from the web. Basic pixel-based skin color classifier mentioned in section 3.1 achieves similar results to those in Jones and Rehg (Jones and Rehg, 2002). We have used RGB bin size = 32 for each channels, and threshold constant $\Theta = 1$. It roughly detects 90% skin color with 14.2% false positive rate.

Superpixel extraction using quick shift is controlled by three parameters: (i) λ controls the trade off between spatial and color consistency, (ii) σ controls the deviation of density estimator, and (iii) τ maximum distance in the quick shift tree. We have used $\sigma = 2$, $\tau = 6$, and $\lambda = 0.9$ for our experiment. Which are chosen using grid search as there is no explicit mechanism to preserve the skin boundaries, with above selected parameters we have noticed that 97.43% skin pixels are correctly grouped into superpixels with 0.35% false positive rate. Average size of the superpixels increases with the larger value of τ and σ and vice versa. Lower values of λ give importance to spatial factor while higher values give importance to the color value. Average size of superpixels are larger when λ is $\cong 0.5$. Skin color detection depends upon the values of the color channels, hence higher importance is given to the color consistency in superpixel extraction. Also, experiments show that the skin boundary is not well preserved with higher spatial importance. The average size of superpixel is 65 in our

(a) Original image (b) Pixel-based (c) Region-based CRF with color information only (d) Region-based CRF with color and border information

Figure 4: Example shows the failure of region-based approach when only a color difference constraint is used with CRF.

experiments. However, the size of superpixels is not fixed and fully depends on the complexity of the images.

Table 1 shows the results comparison between the presented region-based technique and the current state-of-the-art pixel-based skin color detection (Jones and Rehg, 2002) on unconstrained illumination and background. The region based technique without CRF has 91.44% true positive rate with 13.73% false positive rate and with CRF has 91.17% of true positive rate with 13.12% false positive rate. Simply grouping the pixel-based evidence onto superpixels increased the true positive rate by 1.44% and decreased the false positive rate by 0.48%. This shows treating skin as a region yields better results than using pixels only. Both results from the region-based techniques are better than the pixel-based technique.

The results on figure 2 show the effectiveness of the region-based technique with CRF over pixel-based method. Region-based technique first groups the skin and non-skin evidence from each pixels into superpixels level using basic skin color classifier, which helps to remove noise. This is the main reason why only grouping the pixel-based evidence into superpixels increases the true positive rate by 1.44% and reduces the false positive rate by 0.5% (see table 1). Also, CRF helps further extract larger smooth skin regions by exploiting neighbouring color information and boundary sharing between superpixels.

However, there are also some cases where region-based technique performs worse than pixel-based technique when we apply the CRF. Figure 3 are such examples. Skin-like looking pixels and high boundary sharing between skin and non-skin regions are the main reason of the failure. However, we also experimented using the color difference constraint only on CRF instead of both color difference and boundary sharing constraints and found that it performs better when skin regions are very small and narrow. But overall CRF with both neighbour color difference and length of boundary sharing constraints performed better. Figure 4 shows an example where CRF with both neighbours color difference and length of boundary sharing performs better than only with neighbours color difference.

Skin regions do not have the same color values, even the closest skin color pixels within superpixels have different color values. Also, other skin-like objects exist. Hence, results can be further improved using texture information. This is left for our future work.

5 CONCLUSIONS

This paper presents a region-based skin color detection technique, which outperforms the current state-of-the-art pixel-based technique. Color and spatial distance based clustering technique is used to extract the regions from the images, also known as superpixels. In the first step, our technique uses the state-of-the-art non-parametric pixel-based skin color classifier (Jones and Rehg, 2002) which we call the basic skin color classifier. The pixel-based skin color evidence is then aggregated to classify the superpixels. Finally, the CRF is applied to further improve the results. As CRF operates over superpixels, the computational overhead is minimal.

The proposed region-based technique achieved 91.44% true positive rate with 13.73% false positive rate without CRF optimization and 91.17% true positive rate with 13.12% false positive rate with CRF optimization. Grouping the pixel-based evidence into superpixels increased the true positive rate by 1.44% and reduced the false positive rate by 0.48%. Moreover, the region-based approach produced smoother results than the pixel-based methods. Skin commonly appears as regions of similar pixels, so treating skin as a region is advantageous over treating it as an individual pixel. Due to the illumination, background reflection and other noise factors, pixel values vary greatly and grouping them into a region helps to remove noise by collecting evidence from neighbouring pixels.

These results suggest that skin color detection should be region-based rather than pixel-based. Also,

by adding more constraints on the CRF similar to (Shotton et al., 2006) , the detection rate can be improved. Moreover, any better skin color classification method can be used as our basic skin color classification module and can be easily combined with our region-based skin color detection framework defined in section 3 to improve the results.

REFERENCES

Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137.

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1222–1239.

Brown, D., Craw, I., and Lewthwaite, J. (2001). A som based approach to skin detection with application in real time systems. In *Proceedings of the British Machine Vision Conference*, volume 2, pages 491–500.

Fulkerson, B., Vedaldi, A., and Soatto, S. (2009). Class segmentation and object localization with superpixel neighborhoods. In *Proceedings International Conference on Computer Vision*, volume 5.

Jedynak, B., Zheng, H., and Daoudi, M. (2003). Maximum entropy models for skin detection. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 180–193.

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

Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122.

Kawato, S. and Ohya, J. (2002). Automatic skin-color distribution extraction for face detection and tracking. In *International Conference on Signal Processing*, volume 2, pages 1415–1418. IEEE.

Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159.

Kruppa, H., Bauer, M., and Schiele, B. (2002). Skin patch detection in real-world images. In Van Gool, L., editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 109–116. Springer Berlin / Heidelberg.

Moore, A. P., Prince, S., Warrell, J., Mohammed, U., and Jones, G. (2008). Superpixel lattices. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Pan, Z., Healey, G., Prasad, M., and Tromberg, B. (2003). Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560.

Peer, P., Kovac, J., and Solina, F. (2003). Human skin colour clustering for face detection. In *International Conference on Computer as a Tool*.

Phung, S. L., Chai, D., and Bouzerdoum, A. (2002). A universal and robust human skin color model using neural networks. In *Proceedings of International Joint Conference on Neural Networks*, volume 4, pages 2844–2849.

Ren, X. and Malik, J. (2003). Learning a classification model for segmentation. In *IEEE International Conference on Computer Vision*, volume 1.

Sebe, N., Cohen, I., Huang, T., and Gevers, T. (2004). Skin detection: A bayesian network approach. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 903–906, Cambridge, UK.

Shotton, J., Winn, J., Rother, C., and Criminisi, A. (2006). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. *Proceedings of European Conference on Computer Vision*, pages 1–15.

Soatto, S. (2009). Actionable information in vision. In *Proceedings of the International Conference on Computer Vision*, volume 25.

Socolinsky, D. A., Selinger, A., and Neuheisel, J. D. (2003). Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114.

Terrillon, J. C., Fukamachi, H., Akamatsu, S., and Shirazi, M. N. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, page 54.

Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org.

Vedaldi, A. and Soatto, S. (2008). Quick shift and kernel methods for mode seeking. *Proceedings of European Conference on Computer Vision*, pages 705–718.

Wong, K. W., Lam, K. M., and Siu, W. C. (2003). A robust scheme for live detection of human faces in color images. *Signal Processing: Image Communication*, 18(2):103–114.

Yang, M. H. and Ahuja, N. (1998). Detecting human faces in color images. In *International Conference on Image Processing, 1998*, volume 1, pages 127–130.

Zhu, Q., Cheng, K. T., Wu, C. T., and Wu, Y. L. (2004). Adaptive learning of an accurate skin-color model. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 37–42.