Skin Segmentation Using GMM Classifier and Texture Feature Extraction

Chi-Man Pun and Pan Ng

Abstract—In this paper, a skin color segmentation approach by texture feature extraction and k-mean clustering is proposed. We improved the traditional skin classification by combining both color and texture features for skin segmentation. After the color segmentation using a 16 – GMM (Gaussian Mixture Models) classifier, the texture features are extracted using effective wavelet transform with a 2-D Daubechies Wavelet and represented as a list of Shannon entropy. The non-skin regions can be eliminated by the Skin Texture-cluster Elimination using *K*-mean clustering. Experimental results based on common datasets show that our proposed can achieve better performance compared to the existing methods with true positive of 96.5% and with false positives 25.2% for the worst case, with true positive of 90.3% and with false positives 20.5% for the normal case.

Index Terms—Skin segmentation, texture feature, wavelet transform, *k*-mean clustering.

I. INTRODUCTION

Skin Segmentation is principal step in a wide range of image processing applications ranging from sign language recognition [1], human tracking [2], gesture analysis [3], content-based image retrieval systems [4] and to various human computer interaction domains. The target of skin segmentation is to separate the image into skin pixels and non-skin pixels part, skin segmentation should maintain effect performance, whenever the skin region to be rotated, scaled, cut and also partial occluded. Skin Segmentation technique can be categorized as pixel-based and region-based approaches, in pixel-based approaches, only use the color of pixel as a classification feature, on the other hand(region-based approaches), the spatial information of pixels (such as texture) is used to segmentation. The Gaussian mixture models(GMM) [5] is a statistical machine learning method, GMM classifier is more accurately, that use the maximum likelihood estimation to training the classifier, according the classification model, it is made of three main parts: a transducer, feature extractor and classifier. The wavelet transform is a 2-D transform technique, which is excellent as a tool in pattern recognition as well as in numerous other applications. The wavelet transform provides the means of analyzing the input signal into a number of different resolution levels in a hierarchical fashion. This is also known as multiresolution analysis [6]. In this paper, we propose skin color segmentation system that combines color and texture feature of skin to segment skin regions in color images. The present skin color model is extract by GMM

classifier, which is proposed Jones and Rehg [7], although which method is robust to detect the probability of a pixel whether is a skin pixel or not by color feature, he collect a huge amount of web-image as training data of GMM classifier, but it is a pixel-based method, that can't segment any skin color-like object with skin, as shown in Fig. 1, the woman show in under of image, the color of her brown scarf is like as brown skin, so the skin color classifier can't correctly segment as non-skin regions, we propose method used the wavelet transform to extract the texture feature of image, and eliminate the image part that texture not belong to skin.

Fig. 1. The Skin color-like object is detected as skin.

II. RELATED WORKS

The aim of skin segmentation is to isolate skin regions in unconstrained input images, whatever in difficult environment, such as process with a huge amount of the web-images, because of those image is diverse resolution, quality, illumination, noise, etc, and which is a huge and still dataset, implied that can't fetch out the sequential information like as video-base recognition [8], however, traditional skin segmentation methods are not robust enough against with complex environment. The color of skin may vary among different races, age, illumination condition and objects with skin color-like colors, above are major problems must be solve in skin segmentation. The choice of color space can be considered as the primary step in skin segmentation. The skin color model can be considered as the skin color distribution in a color space, and major idea is consider skin color is converge together in a color space and that can be divided by one or several boundary.

The skin color model can be categorized as 2D skin color model and 3D skin color model, the 2D skin color model is observe the skin color in 2D chromatic space [9], the reason is to reduce dependence on lighting variation [10], however,

2D skin color model will cause color feature loss, according several research data, the precision of 3D model always better than that of 2D model, Francesca and Raimondo propose six different algorithms [11] and adopted six color space: YCbCr, RGB, HSV1, HSV2, HIS and rgb. In six algorithms, The YcbCr approach observes skin color as 2D model and other as 3D model, they using Genetic Algorithm (GA) to training the boundary of several color models, their experiment shown that YcbCr color space is more recall oriented and RGB is a good tradeoff between recall and precision [6].

Christophe and Georgios using vector quantization technique [12], to quantize the image color, and construct a 3D skin color model that distributed in HSVand YcbCr color space, both colo models is constructed by 950 skin color samples which have been extracted from various still images and video frames. Jones and Rehg apply visualization techniques [7] to fetch out the shape of skin color distribution. They trained two separate mixture models for the skin and non-skin classes, and used 16 Gaussians in each model, they using 16-GMM classifier with about 3317 training images which formed the non-skin color model and about 1731 training images which formed the skin color model. The advantage of using GMM is that can be made to generalize good result even from small amounts of training data, however it may cause complex computation, slow down performance. Fotouhi propose a method that using nonsubsampled contourlet transforms [13] to represent texture feature of the whole image by row-ordering patch coefficients in each sub-image. They using 300 skin images, 3500 positive and negative training data were extracted and used to train a MLP with two layers of 14 and 9 neurons. Jiang et al employed color, texture and space information to detect skin object in image, and they improved Skin Probability Map (SPM) method, to propose a method that combined by Gabor wavelet transform and Watershed segmentation.
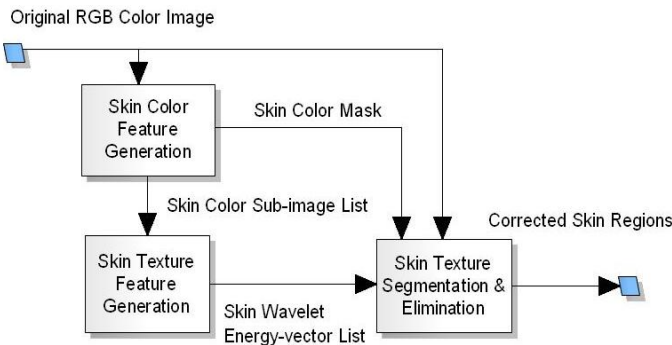

Fig. 2. System modules.

III. SKIN COLOR SEGMENTATION

We propose skin color segmentation system (as shown in Fig. 2) that combined Skin Color Feature Generation, Skin Texture Feature Generation and Skin Texture Segmentation & Elimination three modules, which using both skin color and texture feature to segment skin regions in color images, first original RGB Color image will be inputted into skin color feature generator to generate a Skin Color Sub-image List, and pass into Skin Texture Feature generator to fetch out

Skin Wavelet Energy-vector list, finally Skin Texture Segmentation & Elimination Process will combine all information to generate a Corrected Skin Regions, in the rest parts will give more details procedure in each module.
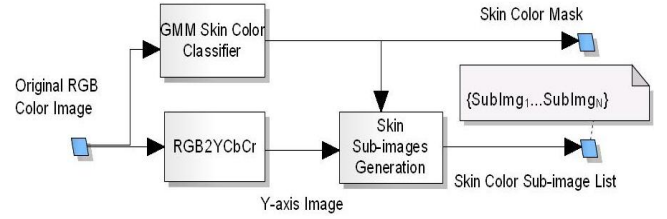

Fig. 3. Skin color feature generation.

A. Skin Color Feature Generation

The Skin Color Feature Generation is a pre-processing of Texture Feature Generation, which isolate the skin and non-skin pixel by using GMM Skin Color Classifier, and build up the Skin Color Mask, but Skin Color Mask only consider color as feature, therefore which may include some skin-color like object, etc. as shown in Fig. 3.

Then Skin Sub-images Generation according skin pixel which is detected by GMM Classifier, for each skin pixel will generate a sub-image to make up the Skin Color Sub-image List, the source of sub-image is a Y-axis images of YCbCr image, which is transformed by original RGB input image, because Y image is essentially a grayscale copy of the main image, which has enough texture information to perform texture segmentation.

1) Gaussian mixture models classifier

The skin color classifier is using 16-GMM Classifier [7] that proposed by Jones and Rehg, the parameters of 16-GMM is same as the result that work out by Jones and Rehg, that is two groups parameters train from skin and non-skin images separately, they are formed by about 3317 non-skin images and 1731 skin images.

A mixture density function is expressed as the sum of Gaussian kernels:

$$P(z) = \sum_{i=1}^{N} W_i \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma_1|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-\mu_i)^T \Sigma_i^{-1}(z-\mu_i)} \quad (1)$$

where z is an original RGB color image and the contribution of the ith Gaussian is determined by a scalar weight w_i, mean vector μ_i, and diagonal covariance matrix Σ_i.

There introduce a concept of skin probability map (SPM), through below formula we can generate a SPM of whole image, calculated by the formula each pixel by pixel.

$$SPM(Z) = \frac{P_{skin(z)}}{P_{nonskin(z)}} \quad (2)$$

The mixture density functions $P_{skin}(z) = P(z)$ where Σ_i of it's using 16 skin diagonal covariance matrix and $P_{nonskin}(z) = P(z)$ where Σ_i of its using 16 non-skin diagonal covariance matrix.

Final, we can generate skin color mask (SCM) from SPM by:

$$SCM(Z) = SPM(Z) \geq TH_{spm} \quad (3)$$

where TH_{spm} is a threshold of SPM, which recommend is set

to 1.

2) Skin sub-images generation

After Skin Color Classifier is a skin color-mask, for each white pixel (x, y) on skin color mask where $CM(x, y) = 1$, the process will cut a square area as sub-image both in skin color mask $CM'(x, y)$ and Y-axis of original image $Y'(x, y)$, they are shown as the grid both in Fig. 4a and Fig. 4b, the equation is shown as equation (3) and equation (4), the result of this process is the Skin Color Sub-image List $\{SubImg_1 \dots SubImg_n\}$.

But this processing will occur some problem is shown as Fig. 6a, the grid include too many background image that is not we major study part, when it case occur, we relinquish to process which pixel, because that will influence the accuracy of result, if we set up the constrain too strict, we will miss too many pixel to be processed, so we introduce the mask-leakage permit concept, if the white pixel on color-mask satisfy equation $(T^2 - size(CM'(x, y))) \leq TH * T^2$ where T is the diameter of sub-mask grid and TH is the maximum mask-leakage permit, default TH always set to 0.2.

About the margin processing details, when fetch the requesting area exceed image's boundary, it will adjust the fetching area within boundary and include $CM(x, y)$ in fetching area.

$$CM'(x, y) = \begin{cases} CM(x', y') \in CM \mid |x'-x| \leq \left[\dfrac{T}{2}\right] \\ \text{and} |y'-y| \leq \left[\dfrac{T}{2}\right] \text{and} CM(x', y') = 1 \end{cases} \quad (4)$$

$$Y'(x, y) = \begin{cases} Y(x', y') \in Y \mid |x'-x| \leq \left[\dfrac{T}{2}\right] \text{and} |y'-y| \leq \left[\dfrac{T}{2}\right] \\ \text{and} \left(T^2 - size(CM'(x, y))\right) \leq TH * T^2 \end{cases} \quad (5)$$
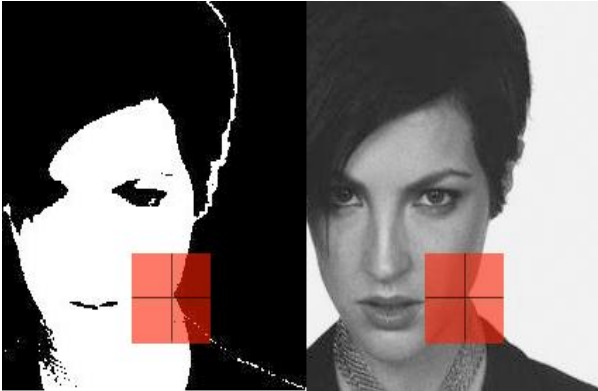
(a) Skin Color Mask   (b) Feting Texture Area
Fig. 4. the process of skin texture filter.

3) Skin color feature generation algorithm

- Input original RGB color image as *z*, using formula (2)

$$SPM(z) = \frac{P_{skin}(z)}{P_{nonskin}(z)} \quad \text{and} \quad \text{formula} \quad (3)$$

$SCM(z) = SPM(z) \geq TH_{spm}$, where CM is Skin Color Mask (SCM) from Skin Probability Map (SPM), CM is a logical matrix, $P_{skin}(z)$ and $P_{nonskin}(z)$ are mixture density functions using 16 skin and nonskin diagonal covariance matrix separately.

- Input original RGB color image as *z*, apply $<Y, Cb, Cr> = RGB2YCbCr(z)$, fetch out *Y* as output which is Y-axis image of *z*.

- Input CM and Y, and then apply formula (4) and formula (5) to generate Skin Color Sub-image List $<Y(z_1), \dots, Y(z_n)>$, where $Y(z_n)$ is a sub-image corresponding to $z_n = (X_n, Y_n)$.
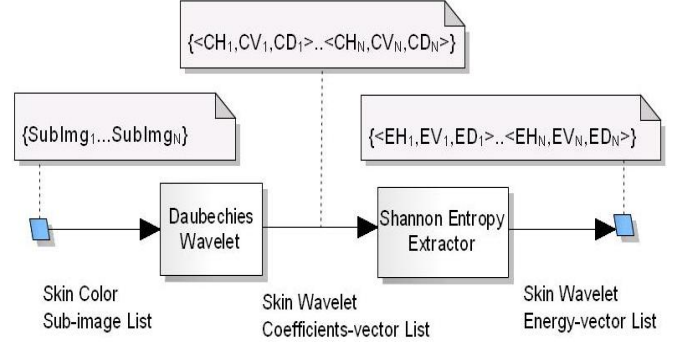

Fig. 5. Skin texture feature generation.

B. Skin Texture Feature Generation

The Skin Color Sub-image List is the product of Skin Sub-images Generation module, for each list item is correspond to a pixel on image which detect as skin pixel by GMM Skin color classifier, the major task of this module is that take a sub-image $SubImg_N$ around these pixel, calculate the 2-D Daubechies Wavelet on these sub-images got Skin Wavelet Coefficients-Vector $<CH_N, CV_N, CD_N>$, then extract the Wavelet Energy-vector $<EH_N, EV_N, ED_N>$ by apply Shannon Entropy on Wavelet Coefficients-Vector, finally the Wavelet Energy-vector will represent the texture feature of these pixel as shown in Fig. 5.

1) 2-D daubechies wavelet

2-D Daubechies wavelet construct by one scaling function $\varphi(x, y)$ (Low-Pass Filter) and three wavelet function $\psi^H(x, y), \psi^V(x, y), \psi^D(x, y)$, (Band-Pass Filter), for scales $j \in Z$, position $n \in Z^2$ and orientation $d \in \{H, V, D\}$

$$\psi_{j^n}^d(x) = \frac{1}{2^j} \psi^d\left(\frac{x - 2^j n}{2^j}\right) \quad (6)$$

where *x* is the input image, and these oriented wavelets are defined by a tensor product of a 1-D wavelet function $\psi(t)$ and a 1-D scaling function $\varphi(t)$

$$\psi^H(X) = \varphi(x_1)\psi(x_2), \psi^v(x) = \psi(x_1)\varphi(x_2) \text{ and } \psi^D(x) = \psi(x_1)\psi(x_2)$$

The scaling function and wavelet function of 1-D Daubechies wavelet is:

$$\psi(X) = \sum_{K=0}^{N-1} a_k \varphi(2x - k) \quad (7)$$

where $(a_0, \dots, a_{N-1})$ is scaling coefficient as finite real series.

$$\psi(X) = \sum_{K=0}^{N-1} b_k \varphi(2x - k) \tag{8}$$

where $(b_0, \ldots, b_{N-1})$ is wavelet coefficient as finite real series.

In this paper we using 1 level 2-D Daubechies wavelet transform with D8 Daubechies coefficients.

2) Shannon entropy

The Shannon entropy equation provides a way to measure the average amount of information contained in random variable, using Shannon entropy represent wavelet energy have several special behavior:

• All coefficients is unsigned value
• The texture is distributed near origin is smooth otherwise is rough.

The Shannon Entropy Equation:

$$E(x) \equiv -\sum_x C(x) \log_2 [C(x)] \tag{9}$$

where $C(x)$ is the distributed data sequence.

The advantage of Shannon entropy is that can observe the distributed energy by a fix min value (origin), that standardization the difference model of observations.

The result of skin sub-images generation is a bunch of sub-image $Y'(x, y)$ that prepare to perform Daubechies wavelet transform, for each sub-images we execute 1- scaling level wavelet transform to decompose them into horizontal, vertical, and diagonal decomposition coefficients vector $< C_{h(x,y)} C_{v(x,y)} C_{d(x,y)} >$, finally each pixel will fetch out a wavelet energy vector(WEV) $< E_{h(x,y)} E_{v(x,y)} E_{d(x,y)} >$ by calculated the Shannon entropy of those decomposition coefficients separately using formula (9).

C. Skin Texture Feature Generation Algorithm

• Input Y' and fetch out each element of that one by one $Y'(z_n), z_n = (x_n, y_n)$, using (6) to (8), calculate wavelet coefficients $C_{h(z_n)} = \psi^H(Y'(z_n))$, $C_{v(z_n)} = \psi^V(Y'(z_n))$, $C_{d(zn)} = \psi^D(Y'(z_n))$, then these coefficients consist a wavelet coefficients vector $< C_{h(z_n)} C_{v(z_n)} C_{d(z_n)} >$. The output is a wavelet coefficients vector list $C = < C_{h(z_1)} C_{v(z_1)} C_{d(z_1)} \ldots C_{h(z_n)} C_{v(z_n)} C_{d(z_n)} >$.

• Input C and fetch out each element of that one by one $C(z_n), z_n = (x_n, y_n)$, using (9) to calculate wavelet energy vector along each component,
$$E_a(z_n) \equiv -\sum_x C_a(z_n) \log_2 \log_2 [C_a(z_n)], a \in \{h, v, d\}$$
, the output is a wavelet energy vector list $E = < E_{h(z_1)} E_{v(z_1)} E_{d(z_1)} \ldots E_{h(z_n)} E_{v(z_n)} E_{d(z_n)} >$.

1) Skin texture segmentation and elimination

The result of Skin Texture Generation is a list of Skin Wavelet Energy-vector $< EH, EV, ED >$, these vectors is the observation data of k-mean clustering, and then these observation data will be separated into several clusters $\{Cluster_1 \ldots Cluster_N\}$, among these clusters, some clusters is non-skin regions on original image, which will be eliminated by Skin Texture-cluster Elimination, finally Skin Regions Generation will combine RGB input image, Skin color mask and Corrected Skin Texture-segment cluster list,

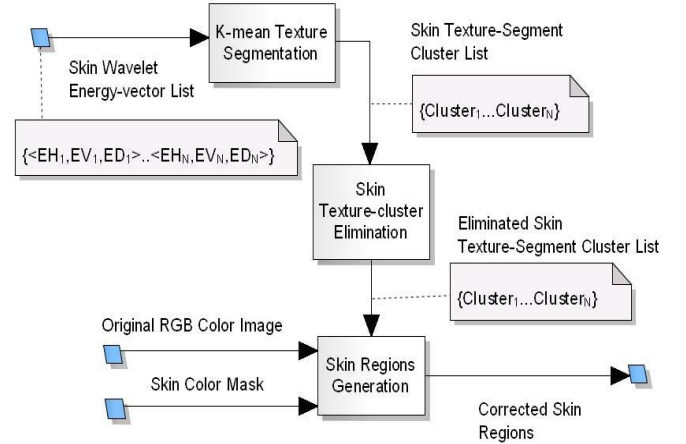to regenerate a corrected skin regions images as shown Fig. 6.

Fig. 6. Skin Texture Segmentation & Elimination.

2) K-means clustering

Given a set of observations $(x_1, \ldots, x_n)$, where each observation is a d-dimensional real vector, the aim of k-means clustering is an algorithm to cluster n observations into k partitions $(k < n)$, and tries to achieve is to minimize total intra-cluster variance.

$$arg_s min \sum_{i=1}^{k} \sum_{X_j \in S_i} \| X_j - \mu_i \|^2 \tag{10}$$

where there are k clusters S_i, $i = 1, 2 \ldots, k$, and μ_i is the centric of all the points x_j in S_i.

The Shannon entropy extract a serious of WEV $< E_{h(x,y)} E_{v(x,y)} E_{d(x,y)} >$, there vector will be observed as distributed in a 3-dimension space, and then apply k-mean clustering to segment the observation data into K cluster set $S = \{S_1, S_2, \ldots, S_k\}$ and for each cluster which correspond to a centric point μ_i, there construct a centric point list $\mu = \{\mu_1, \mu_{2s}, \ldots, \mu_n\}$ as shown in Fig. 7.

3) Skin texture-cluster elimination

The basic idea of skin texture segmentation is to elimination rough part of WEV, according the property of Shannon entropy, the WEV near origin means that is smoother otherwise is rougher, then we remain R cluster which mean nearer to origin and eliminate other part, then non-skin parts will be eliminated.

$$\mu' = ASCSort(\mu), S' = \{S_i \subseteq S | \mu_i > \mu'_r\} \tag{11}$$

where μ' is the centric point list which is sorted by increasing order, r is the number of remain cluster, S' is the cluster sets that should be eliminated as shown in Fig. 8.

4) Skin regions generation

Finally, we got a list of eliminated WEV cluster, actually each WEV correspond to a pixel on original images, and then we fetch out these pixels and eliminate them from skin color mask to gain a corrected skin regions mask, and apply it mask to original color image, to fetch out correct skin regions image.

For example the original image in Fig. 9(a), by using 16-GMM skin color classifier, it will generate the result as Fig. 9(b), but the brown scarf is detected as skin, after using Daubechies wavelet and using 16×16 sub-mask generate the wavelet energy distribution is shown in Fig. 7, then we apply

5-mean (K=5) clustering and remain 1(R=1) cluster which mean nearest the origin, finally got the fixed skin regions image Fig. 9(c).
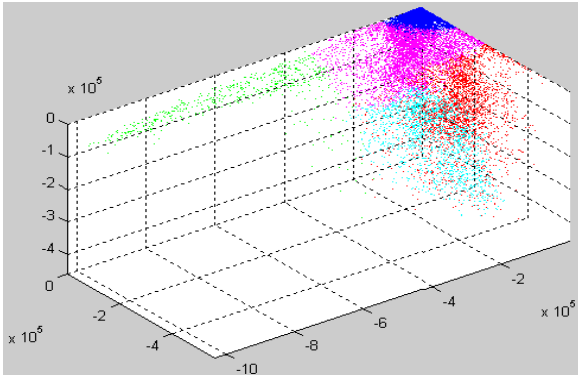

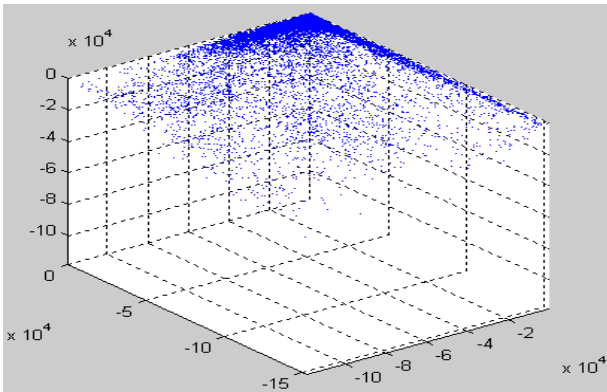Fig. 7. The *k*-mean clustering of WEV of the original image in Fig. 1.

Fig. 8. The remained clustering of Fig. 9. after skin texture-cluster elimination.

About the parameter of sub-mask size *T*, *K* value of *k*-mean clustering and remain cluster num *R*, that can be train out by other Neural Networks method, normally, we recommend that $T = 16$, $K = 6$ and $R = 1$.
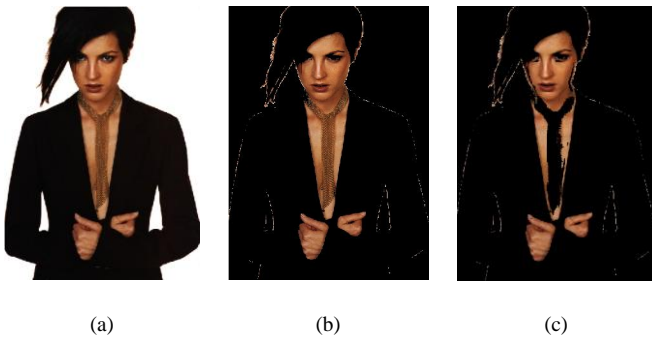

(a)   (b)   (c)

Fig. 9. Skin segmentation process snapshot. (a) original images, (b) after 16-GMM skin color classifier brown scarf was detected as skin, (c) after skin feature filter with 16×16 sub-mask and 5-mean clustering remain 1 minimum cluster, then the brown scarf been eliminated.

IV. EXPERIMENTAL RESULTS

For the experiments, we employed two group of testing images from the common dataset as in [7]. The first group is the 200 worst case images, and totally 41261256 pixels are detected as skin pixel by using GMM classifier, but there exists a lot of skin color-like object with skin, the precision of result with gain high false positive rate (FPR) by 45.7% The

second group is 200 skin images by random selection, and totally 39211281 pixels are detected as skin pixels. As shown in Table I, we have improved the GMM classifier method successfully, in worst case, our method can detect the skin-like object and eliminate it, although true positive rate (TPR) is slightly decreased to 96.5%, but we also decrease FPR from 45.7% to 25.2%. In normal case, we have improved the performance for both true positive and false positive, compared with the original GMM classifier. This is mainly due to our proposed method has integrated the texture features with color features for the skin detection, so that some skin regions and non-skin regions can be effectively separated.

V. CONCLUSION

Skin segmentation is the basic of many important image processing systems. In this paper, we successfully improve the precision of previous GMM algorithm, which just considers the color feature as unique segmentation feature, caused a lot of skin color-like object detect as skin, that increase the FPR. We proposed a skin color segmentation system which combine color and textures feature as segmentation feature, eliminate non-skin part that detect as skin by previous method, finally we got the average detection rate of our method by 90.3% true positive with 20.5% false positive. Future work may focus on decrease the computational overhead.

TABLE I: TRUE AND FALSE POSITIVE RATE OF THE PROPOSED METHOD

Worst case		
GMM classifier	true positive	false positive
	97.8%	45.7%
Our Method	true positive	false positive
	96.5%	25.2%

Normal case		
GMM classifier	true positive	false positive
	88.7%	24.7%
Our Method	true positive	false positive
	90.3%	20.5%

REFERENCES

[1] J. Han, G. Awad *et al*., "Automatic skin segmentation and tracking in sign language recognition," *Computer Vision, IET*, vol. 3, no. 1, pp. 24-35, 2009.

[2] K. Bhuvaneswari and H. A. Rauf, "Edgelet based human detection and tracking by combined segmentation and soft decision," in *Proc. International Conference on Control, Automation, Communication and Energy Conservation*, 2009.

[3] K. K. Kyung, K. K. Chang *et al*., "Gesture analysis for human-robot interaction," in *Proc. The 8th International Conference Advanced Communication Technology*, 2006.

[4] K. H. Yap and K. Wu, "A soft relevance framework in content based image retrieval systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1557–1568, December 2005.

[5] H. Permuter, J. Franco *et al*., "Gaussian mixture models of texture and colour for image database retrieval," in *Proc. IEEE International Conference on Acoustics*, *Speech, and Signal Processing*, 2003.

[6] C. Zhang, H. X. Zheng *et al*., "3-D wavelet-based multiresolution analysis," in *Proc. 5th International Symposium on Antennas, Propagation and EM Theory*, 2000.

[7] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[8] D. Yingzi, E. Arslanturk *et al*., "Video-based noncooperative iris image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 64-74, 2011.

[9] J. Kovac, P. Peer et al., "Human skin color clustering for face detection," *The IEEE Region 8, Computer as a Tool*, 2003.

[10] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Science Direct. Pattern Recognition*, vol. 40, issue 3, pp. 1106-1122, March 2007.

[11] F. Gasparini and R. Schettini, "Skin segmentation using multiple thresholding," in *Proc. SPIE 6061*, 2006.

[12] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264-277, 1999.

[13] M. Fotouhi, M. H. Rohban et al., "Skin detection using contourlet texture analysis," in *Proc. 14th International CSI Computer Conference*, 2009.

Chi-Man Pun received his B.Sc. and M.Sc. degrees in software engineering from the University of Macau in 1995 and 1998 respectively, and Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong in 2002. He is currently an associate professor at the Department of Computer and Information Science of the University of Macau. He has investigated several funded research projects and published more than ninety refereed scientific papers in international journals, books and conference proceedings. Dr. Pun has also been invited to serve as general/program and session chairs for several international conferences, reviewer and/or committee member for many international journals. His research interests include digital image processing; digital watermarking; multimedia security; pattern recognition and computer vision; intelligent systems and applications. He is also a senior member of the IEEE and a professional member of the ACM.