Skin Detection Based on Image Color Segmentation with Histogram and *K*-Means Clustering

Emir Buza, Amila Akagic, Samir Omanovic

University of Sarajevo
Faculty of Electrical Engineering, Department for Computer Science and Informatics
Zmaja od Bosne bb, Kampus Univerziteta, 71000 Sarajevo
emir.buza, amila.akagic, samir.omanovic@etf.unsa.ba

Abstract

Skin detection is a crucial pre-processing step for finding human faces in images. The challenging task is to find a reliable, yet efficient method for detection of skin region(s). In this paper, we proposed a new, simple and efficient method for skin detection based on image segmentation of different color spaces, and simple clustering technique (K-means) for clustering similar pixels on an image. Three K-means input features are used : a) two components from two different color spaces (Hue, Cr, Cb), b) positions of pixels on an image and c) rough estimation of skin pixels obtained from skin-color based detection. Our approach showed promising results on human images from different ethnicities, with simple background and high illumination. The computational cost of the method has been very low, since no training data is required. Results indicate that the method is suitable as a pre-processing step for some supervised method for advanced human skin segmentation and detection.

Keywords: Skin detection, Unsupervised method, *k*-means clustering, Image processing, Image segmentation

I. Introduction

A reliable and efficient human skin detection has been the first necessary step in many image processing applications, such as face detection and tracking, gesture analysis, content-based image retrieval systems, de-identification, privacy-protection and other human computer interaction domains. In recent years, numerous skin detection methods have been presented in literature. They vary from methods based on manipulation of color-space channels to more sophisticated statistical modeling and machine learning methods. The former have been the most common methods in literature, and they are in general considered as computationally effective. The common steps in a skin detection algorithm usually include transformation of skin pixels into an appropriate color space and classification through labeling of skin pixels into skin and non-skin pixels (skin classifier). The common problem has usually been high false skin detection, which must be corrected with some additional method(s). The latter methods required a pre-processing step for training a binary classification system, and have one major drawback: the classifier performance highly depends on the size of a training set. Thus, the existing solutions usually make a trade-off between precision and computational complexity.

Detection of human skin has been a challenging task because many factors affect skin appearance in images [1] [2]. These factors include illumination, camera characteristics, ethnicity, individual characteristics, background characteristics, etc. There are three main problems when designing a method based on skin color as a feature. These are: what color space to choose? how to model skin and non-skin pixel distribution? and how to classify the modeled distribution? This paper attempts to provide answers to all three problems. There are several color spaces with different properties in literature. The most popular color spaces are RGB, Normalized RGB, HSV, TSL and YCrCb [3]. The RGB color space is the default color space for the most image formats, while other color spaces can be obtained with linear and non-linear transformation of RGB color space. The choice of color space significantly influences modeling efficiency of the skin-color distribution. The goal of skin modeling is to build a decision rule in order to classify skin and non-skin pixels. There have been several modeling choices: explicitly defined skin region (a number of rules), non-parametric (histogram-based), parametric (Gaussian, Elliptic boundary model) and dynamic skin distribution (used for face tracking). A comprehensive survey of skin modeling distributions has been published in [2], [4].

In this paper, a new region-based method for skin detection based on skin-color information has been presented. The primary steps for skin detection in our method include (1) combining two color spaces, HSV and YCrCb, to represent image pixels, (2) using histogram-based modeling with new approach for defining threshold value and (3) using K-means clustering for unsupervised classification with new data set for pre-processing. K-means clustering [5] is unsupervised, non-deterministic technique for generating a number of disjoint and flat (non-hierarchical) clusters. It is used to cluster similar pixels with an equal cardinality constraint. Special dataset consisting of three input features has been defined for clustering image pixels into three clusters: background, foreground and skin pixels.

Our method is simple and fast when compared to the existing state-of-the-art segmentation methods for skin detection. It has low computational cost, since it does not require pre-processing of large training dataset. The experimental results showed that changes in illumination conditions and viewing environment do not affect quality of skin detection. Results also indicated that the method has no limitation on choice of ethnicity. On complex backgrounds, the method detected objects with similar color as skin, thus some improvements are necessary for distinguishing background objects with similar colors. The method produced better results for images with high level of skin area, while on images with many small regions (i.e. multiple humans on a image) the results are modest. However, the method is suitable as a pre-processing step for some sophisticated supervised method.

The paper is organized as follows: related work on skin detection using image processing is briefly reviewed in Section II. In Section III, we proposed a new method for skin detection. Implementation and experimental results have been presented in Section IV, and in Section V we concluded the paper with some remarks.

II. Related Work

Three comprehensive surveys of skin-color modeling and detection methods have been presented by Kakumanu et al. [2], Mitra et al. [6] and Vezhnevets et al. [4]. The process of skin detection usually consists of three steps: transformation of image pixels into an appropriate color space, selection of appropriate distribution and classification of skin or non-skin pixels by some classifier with some labeling procedure. Existing solutions have usually been making a trade-off between precision and computational complexity. The latter is usually extensive in supervised methods, where it is advisable to process a comprehensive training data set. This has been negatively affecting the execution time, even though it has been shown that the accuracy of the algorithm is enhanced. An example has been given in Jones and Rehg [7], who used the Bayesian decision rule with a 3-D RGB histogram model built from 2 billion pixels collected from 18696 web images to perform skin detection. Methods [8], [9] and [10] used fixed range values for color space, however the drawback was high false skin condition in cases with wide variety of skin colors from different ethnicities, high illumination or complex background. Other approaches include the use of luminance invariant colors spaces [11], [12], neural networks [13], Bayesian classifiers [14] and random forest [15].

It is challenging to compare skin detection methods precisely and fairly, since there is no available standard nor unified dataset. Different methods either use different existing datasets or some self-generated datasets, which are not publicly available. One problem of existing datasets has been the exact definition of skin pixel (i.e. can human lip be considered as skin pixel?), which has not been determined identically among different datasets and authors [16]. The other problem has been related to ground truth images, which are often inaccurate due to human error or semi-automatic procedure. The effect is misinterpretation of results.

III. Our method

Fig. 1 shows the proposed method for skin detection. First, an image is segmented to detect relevant information (Step 1). Second, background pixels with defined threshold are removed (Step 2). Third, skin pixels are detected (Step 3) with two color spaces (HSV and YCrCb). This step provided rough estimation of skin pixels, thus a new refinement stage was necessary. A new dataset was defined (Step 4) in order to cluster pixels on an image. In the Step 5, pixels were clustered into three clusters: background, foreground and skin pixels with unsupervised *K*-means clustering.

III-A. Image segmentation

Image segmentation is usually the first step of image analysis and pattern recognition. There are several methods to perform image segmentation including thresholding, clustering, transform and texture methods [17] [18] [19]. The simplest approach to segment an image is histogram-based thresholding, which we also used. The method assumes that an image is composed of different colors or gray regions. The first step was conversion of pixel values from RGB into gray scale. The pixels are partitioned depending on their intensity values and results should reveal two peaks, corresponding to background and foreground object(s). Histogram-based thresholding consists of setting intensity values - threshold (η), which separates background from the object(s) of interest.

In the Step 1 (Fig. 1), we determined a new threshold value η (Eq. 1) with Otsu's image thresholding method [17] [20] and maximum color frequency of histogram $h(x, y)$. The results are shown in Fig. 2. A red line on both graphs represents the threshold η. This value is calculated based on two values: (1)

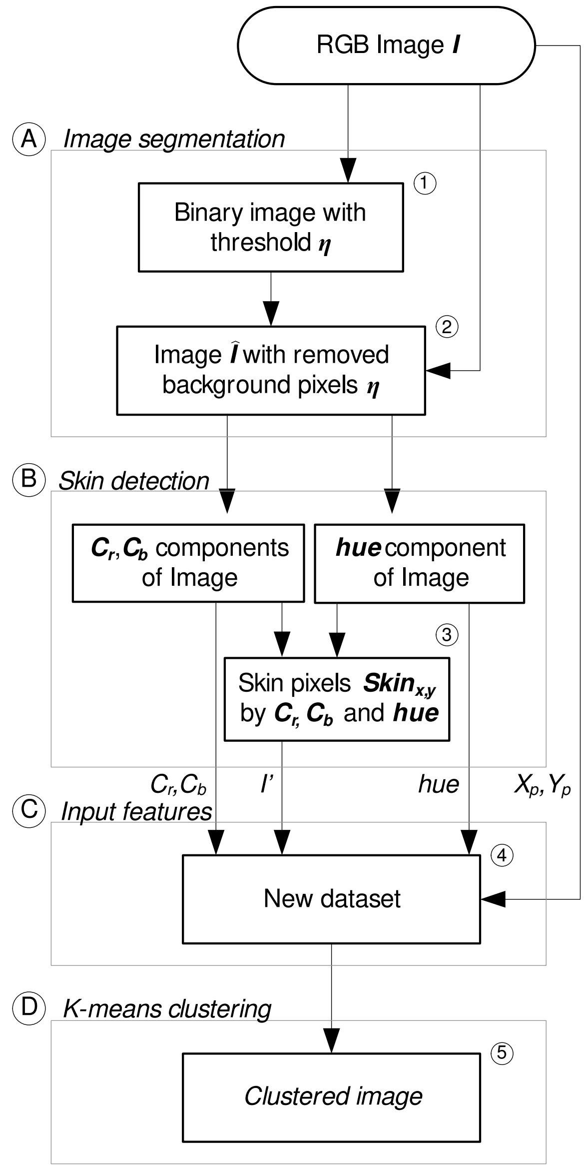


Fig. 1. Proposed method for skin detection on an image.

T_{Otsu} (green point 'o'), which is the threshold provided by Otsu's method, and (2) T_{max} (blue point 'o'), which is the maximum point of color frequency on histogram $h(x, y)$.

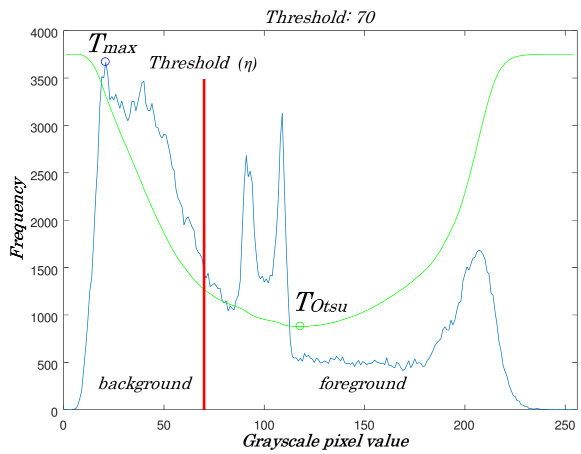
$$\eta = \begin{cases} round(\frac{T_{Otsu}+T_{max}}{4}), & \text{if } T_{max} \leq 10 \\ round(\frac{T_{Otsu}+T_{max}}{2}, & \text{if } T_{max} > 10 \end{cases} \quad (1)$$

When $T_{max} \leq 220$ (Fig. 2a), the image I is converted from the histogram $h(x, y)$ to a binary image $g(x, y)$ using Eq. 2, otherwise we use Eg. 3 (Fig. 2b) .

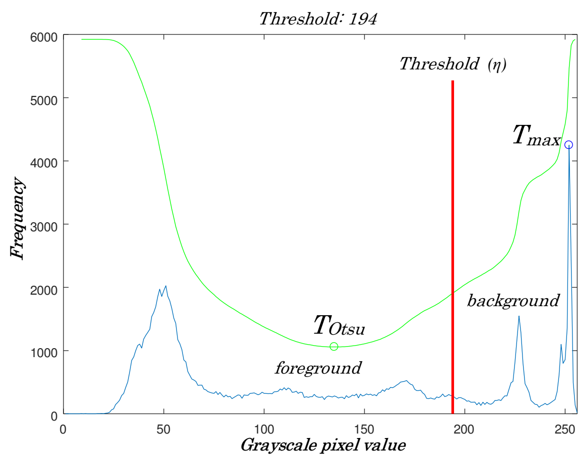
$$g(x, y) = \begin{cases} 0, & \text{if } h(x, y) \leq \eta \\ 1, & \text{if } h(x, y) > \eta \end{cases} \quad (2)$$

$$g(x, y) = \begin{cases} 0, & \text{if } h(x, y) \geq \eta \\ 1, & \text{if } h(x, y) < \eta \end{cases} \quad (3)$$

In the Step 2 (Fig. 1), background pixels are removed from the original image I based on information from the



(a) Background is on the left side from the red line, and foreground is on the right.



(b) Foreground is on the left side from the red line, and background is on the right.

Fig. 2. The process of determining threshold η.

binary image $g(x, y)$. This operation is executed for each color component of an image (Eq. 4).

$$\hat{I}(R, G, B) = \begin{cases} (0, 0, 0), & \text{if } g(x, y) = 0 \\ (I_{R(x,y)}, I_{G(x,y)}, I_{B(x,y)}), & \text{if } g(x, y) = 1 \end{cases} \tag{4}$$

III-B. Skin detection

RGB color space [21] is considered as the default color space for the most image formats, since it is more sensitive to different light conditions. However, better results for skin color classification might be achieved with other types of color spaces, for instance HSV and YCrCb color spaces. These color spaces are less sensitive to light conditions. Hue-saturation based color spaces describe color with intuitive values on the basis of artist's idea of tint, saturation and tone. Hue defines the dominant color of an area, while saturation measures the colorfulness of an area in proportion to its brightness. Hue has an interesting property: it is invariant to highlights at white light sources. YCrCb color space represents color with luma (luminance, nonlinear RGB transformation) and two color difference values Cr and Cb formed by subtracting luma from RGB red and blue components [4].

In the Step 3 (Fig. 1), we first converted an image from RGB to HSV color space. Then, we calculated Cr and Cb components with Eq. 5.

$$\begin{aligned} Cr_{x,y} &= 0.439 * \hat{I}_{R(x,y)} - 0.368 * \hat{I}_{G(x,y)} - 0.071 * \hat{I}_{B(x,y)} \\ &\quad +128 \\ Cb_{x,y} &= 0.148 * \hat{I}_{R(x,y)} - 0.291 * \hat{I}_{G(x,y)} + 0.439 * \hat{I}_{B(x,y)} \\ &\quad +128 \end{aligned} \tag{5}$$

Potential skin color pixels are selected with Eq. 6.

$$Skin_{x,y} = \begin{cases} 1, & \text{if } Cr_{x,y} \geq 140 \text{ and } Cr_{x,y} \leq 160 \\ & \text{and } Cb_{x,y} \geq 140 \text{ and } Cb_{x,y} \geq 170 \\ & \text{and } hue_{x,y} \leq 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

The final result of the Step 3 is an image I'. This image has been formed from $Skin_{x,y}$ by converting pixel values "1" into value "255", while "0" values are unchanged.

III-C. Input Features

In the Step 4, we defined special dataset made of input features in order to cluster pixels on an image. This dataset contains (1) some components of two color spaces (Hue, Cr, Cb), (2) positions of pixels on the image (X_p, Y_p) and (3) rough estimation of skin pixels (I') obtained from the Step 3. Since all information is contained in a matrix, we converted all six aforementioned components (Cr, Cb, Hue, X_p, Y_p and I') into appropriate vectors. The result is a new dataset with six columns and $N \times M$ rows, where N is a number of rows and M is a number of columns in a matrix.

III-D. *K*-means Clustering

K-means clustering is a simple method which uses distance measure for grouping data into *K* pre-defined number of groups (clusters). In our case, image pixels are clustered into three clusters: background, foreground and skin pixels. We used square Euclidean measure as a distance. Skin pixels defined in the Step 3 determine which clusters represents skin (in our case, white color represents skin on clustered image).

IV. Implementation and Results

IV-A. Implementation

Our method has been implemented in MATLAB version 7.11.0 (R2010b) with Image Processing Toolbox. The image processing was performed on Intel Core2 2.80 GHz CPU with 8GB of RAM.

The effectiveness of our method has been verified on a data collection selected from the Pratheepan dataset[1]. The Pratheepan dataset consists of a set of images downloaded randomly from Google. This approach allows convenient verification and comparison with similar methods.

IV-B. Results

Detection capabilities of our method are shown in Fig. 3 and Fig. 4, and a quantitative analysis on the Pratheepan dataset is presented in Table I. The *accuracy* is a score that uses the sum of *sensitivity* (true positive - skin pixels) and *specificity* (true negative - non skin pixels). F-score is calculated using measure of *precision* and *recall*. Column *Avg* provides average value for all images in two sets: Face (Fig. 3) and Family Photos (Fig. 4).

[1]The dataset is available at http://www.cs-chan.com/downloads_skin_dataset.html

	Face Photos	Family Photos	Avg
Sensitivity	0.70	0.55	0.63
Specificity	0.95	0.78	0.87
Precision	0.70	0.55	0.63
Recall	0.84	0.32	0.58
Accuracy	0.88	0.74	0.81
F-Score	0.74	0.39	0.57

TABLE I. A quantitative analysis of our method on the Pratheepan dataset.

The method shows promising results on human images from different ethnicities, with simple background and high illumination. The results on images with high level of skin area are better than on images with low level of skin. However, on some complex backgrounds (Fig. 4: image 5, 6 and 7) the method detects objects with similar color as skin even though it is not skin. Some improvements are necessary for detection of multiple humans on an image or for smaller areas of skin pixels. It has low computational cost when compared to supervised methods, since it does not require any training stage,

Table I shows that proposed method has acceptable score when compared to [1]. The method has better results for the Face Photos, while the Family Photos results are modest. The best result is achieved for the case when images contain a lot of skin pixels with simple backgrounds. In that case, the method has very small error (Specificity is 0.95).

V. Conclusion

In this paper, a new region-based skin detection method based on image color segmentation with histogram and K-means clustering has been presented. An image was segmented with histogram based data from gray scale image and a new image without background pixels was created. Two different color spaces were used for skin color detection. The components of these color spaces were used to create a new dataset to cluster pixels on an image with simple K-means clustering.

The proposed method was tested on *Face and Family photos* from the Pratheepan dataset. The images were selected based on different conditions, such as ethnicity, background and illumination. The method is suitable for skin detection and has low computational cost. The results indicated that the method detects skin color with reasonable accuracy and it is suitable for image pre-processing. The future work will focus on better detection of multiple humans and smaller areas of skin pixels on an image.

VI. References

[1] Tan, W. R., Chan, C. S., Yogarajah, P., & Condell, J. (2012). *A fusion approach for efficient human skin detection.* IEEE Transactions on Industrial Informatics, 8(1), pp. 138-147.

[2] Kakumanu, P., Makrogiannis, S., & Bourbakis, N. (2007). *A survey of skin-color modeling and detection methods.* Pattern recognition, 40(3), pp. 1106-1122.

[3] Mendenhall, M. J., Nunez, A. S., & Martin, R. K. (2015). *Human skin detection in the visible and near infrared.* Applied optics, 54(35), pp. 10559-10570.

[4] Vezhnevets, V., Sazonov, V., & Andreeva, A. (2003, September). *A survey on pixel-based skin color detection techniques.* In Proc. Graphicon (Vol. 3, pp. 85-92).

[5] Jain, A. K. (2010). *Data clustering: 50 years beyond K-means.* Pattern recognition letters, 31(8), pp. 651-666.

[6] Mitra, S., & Acharya, T. (2007). *Gesture recognition: A survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(3), pp. 311-324.

[7] Jones, M. J., & Rehg, J. M. (2002). *Statistical color models with application to skin detection.* International Journal of Computer Vision, 46(1), pp. 81-96.

[8] Sobottka, K., & Pitas, I. (1998). *A novel method for automatic face segmentation, facial feature extraction and tracking.* Signal processing: Image communication, 12(3), pp. 263-281.

[9] Wang, Y., & Yuan, B. (2001). *A novel approach for human face detection from color images under complex background.* Pattern Recognition, 34(10), pp. 1983-1992.

[10] Omanovic, S., Buza, E., & Besic, I. (2014, May). *RGB ratios based skin detection.* In Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on (pp. 1348-1353). IEEE.

[11] Yang, U., Kim, B., & Sohn, K. (2009, May). *Illumination invariant skin color segmentation.* In Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on (pp. 636-641). IEEE.

[12] Vadakkepat, P., Lim, P., De Silva, L. C., Jing, L., & Ling, L. L. (2008). *Multimodal approach to human-face detection and tracking.* IEEE transactions on industrial electronics, 55(3), pp. 1385-1393.

[13] Mohamed, A., Weng, Y., Jiang, J., & Ipson, S. (2008, July). *Face detection based neural networks using robust skin color segmentation.* In Systems, Signals and Devices, 2008. IEEE SSD 2008. 5th International Multi-Conference on (pp. 1-5). IEEE.

[14] Sebe, N., Cohen, I., Huang, T. S., & Gevers, T. (2004, August). *Skin detection: A bayesian network approach.* In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on (Vol. 2, pp. 903-906). IEEE.

[15] Khan, R., Hanbury, A., & Stoettinger, J. (2010, September). *Skin detection: A random forest approach.* In Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 4613-4616). IEEE.

[16] Mahmoodi, M. R., Sayedi, S. M., Karimi, F., Fahimi, Z., Rezai, V., & Mannani, Z. (2015, November). *SDD: A skin detection dataset for training and assessment of human skin classifiers.* In Knowledge-Based Engineering and Innovation (KBEI), 2015 2nd International Conference on (pp. 71-77). IEEE.

[17] Parker, J. R. (2010). *Algorithms for image processing and computer vision.* John Wiley & Sons.

[18] Mignotte, M. (2008). *Segmentation by fusion of histogram-based k-means clusters in different color spaces.* IEEE Transactions on image processing, 17(5), pp. 780-787.

[19] *Matlab Toolbox for Image Processing, Image Analysis, Region and Image Properties http://www.mathworks.co.uk/help/images/ref/regionprops.html#bqkf8id.*

[20] Zhang, J., & Hu, J. (2008, December). *Image segmentation based on 2D Otsu method with histogram analysis.* In Computer Science and Software Engineering, 2008 International Conference on (Vol. 6, pp. 105-108). IEEE.

[21] Kumar, A. (2014, February). *An empirical study of selection of the appropriate color space for skin detection: A case of face detection in color images.* In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on (pp. 725-730). IEEE.

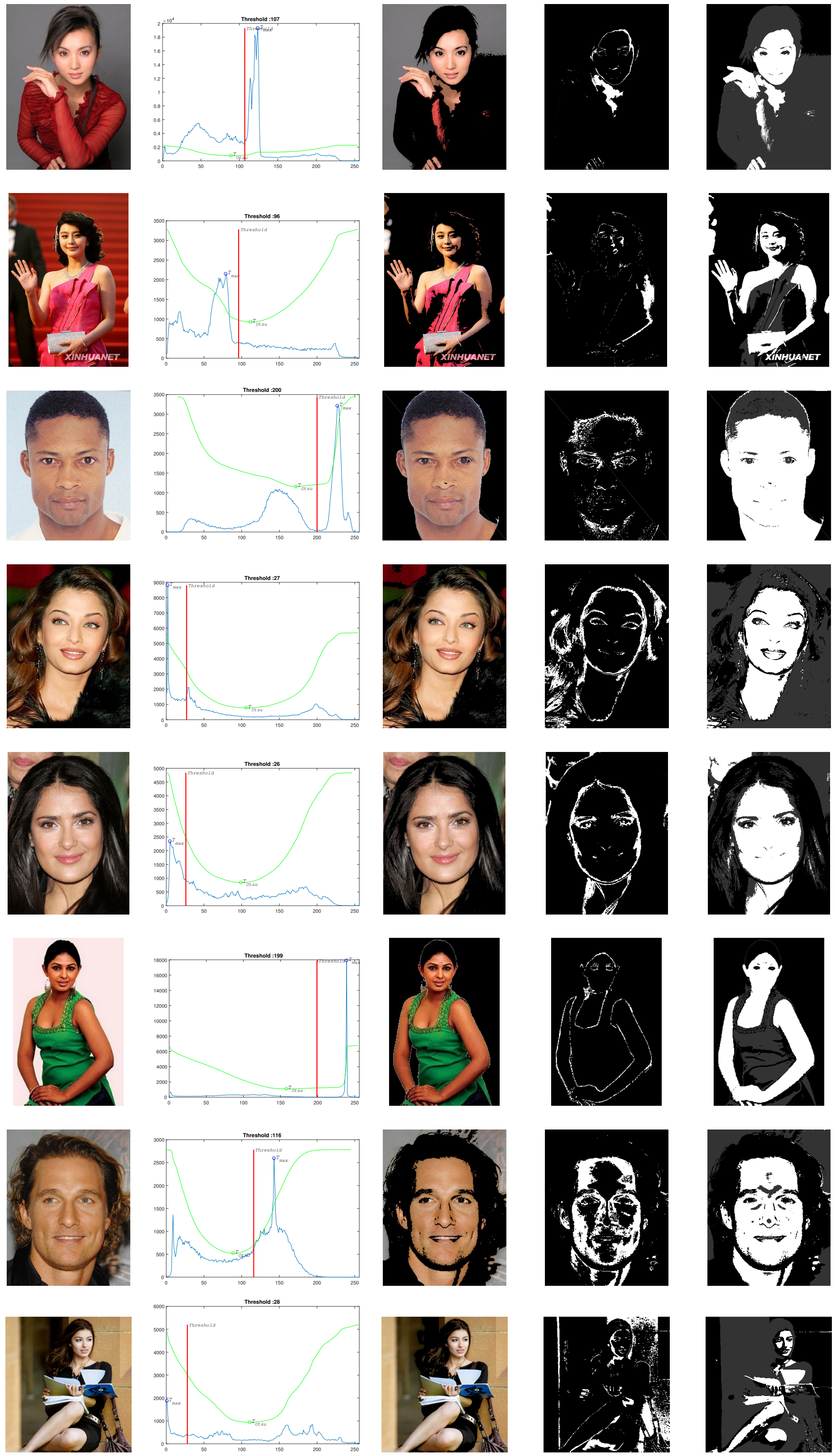


Fig. 3. Detection capabilities of our method with single human on an image and simple background: (1) original images, (2) histogram with threshold, (3) image with removed background pixels, (4) detected skin pixels and (5) clustered skin color detection.

Fig. 4. Detection capabilities of our method with multiple humans on an image and complex background: (1) original images, (2) histogram with threshold, (3) image with removed background pixels, (4) detected skin pixels and (5) clustered skin color detection.