Applied Soft Computing 33 (2015) 337–347

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Hybrid Human Skin Detection Using Neural Network and K-Means Clustering Technique

Hani K. Al-Mohair, Junita Mohamad-Saleh*, Shahrel Azmin Suandi

School of Electrical & Electronic Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia

ARTICLE INFO

Article history:
Received 5 December 2014
Received in revised form 19 April 2015
Accepted 26 April 2015
Available online 7 May 2015

Keywords:
Skin color detection
Color space
Neural networks
Texture analysis
k-Means

ABSTRACT

Human skin detection is an essential step in most human detection applications, such as face detection. The performance of any skin detection system depends on assessment of two components: feature extraction and detection method. Skin color is a robust cue used for human skin detection. However, the performance of color-based detection methods is constrained by the overlapping color spaces of skin and non-skin pixels. To increase the accuracy of skin detection, texture features can be exploited as additional cues. In this paper, we propose a hybrid skin detection method based on YIQ color space and the statistical features of skin. A Multilayer Perceptron artificial neural network, which is a universal classifier, is combined with the k-means clustering method to accurately detect skin. The experimental results show that the proposed method can achieve high accuracy with an F_1-measure of 87.82% based on images from the ECU database.

1. Introduction

Separating an image into regions that consist of groups of identical linked pixels is an image processing stage called image segmentation. The homogeneity of a region can be defined by the color, gray levels, and texture, among other factors [1]. Skin detection is a good example of image segmentation, which can be accomplished by classifying image pixels as skin and non-skin pixels [2].

The importance of skin color detection comes from its use as a primary operation in many applications, such as face detection [3], surveillance systems [4], Internet pornographic image filtering [5] and gesture analysis [6]. For example, face detection is accomplished by taking out the joint facial characteristics and by employing skin color detection as a primary step to specify the face area. As a result, accurate and fast face detection can be accomplished.

Many researchers have investigated skin color features. Their results have shown that skin color has a limited range of hues and is not deeply saturated [7]. Thus, human skin color is clustered within a small area in the color space. In the past few years, several algorithms have been proposed for skin detection. However, some factors, such as illumination, may make skin color detection a very difficult task. [2]. The existing algorithms can be categorized into four classes: explicit skin classifiers, parametric classifiers, nonparametric classifiers, and dynamic classifiers [8]. The explicit classifiers, which are the easiest and are frequently employed, use a threshold strategy to distinguish between skin and non-skin pixels [9]. Basically, they characterize the limits of the skin region by utilizing a set of fixed thresholds. Although such classifiers are direct and might be used without any prior training steps, they may need adaptability when utilized under distinct imaging conditions. This may result in incorrect pixel detection [8].

Parametric classifiers can be based on a single Gaussian model [10], multiple Gaussian clusters [11], a mixture of Gaussian (MoG) models [12], or an elliptic boundary model [13]. Generally, the characterization speed of these classifiers is slow. In fact, they need to process each pixel individually. Additionally, these methods have low detection accuracy, as they rely on approximated parameters rather than authentic appropriate skin colors [8]. Furthermore, their performance varies depending on the utilized color space [14].

In the nonparametric classifiers, a set of training data is essential for estimating the statistical model of skin color distribution [15]. The advantages of these classifiers are quick training and skin distribution shape independence [14]. Nevertheless, such classifiers are not precise enough because of the requirement for an unbounded amount of training information, which makes them appropriate in a constrained scope of imaging conditions [8].

* Corresponding author. Tel.: +60 45996027.
E-mail addresses: hamo_17@yahoo.com (H.K. Al-Mohair), jms@usm.my (J. Mohamad-Saleh).

http://dx.doi.org/10.1016/j.asoc.2015.04.046
1568-4946/

To overcome the generality of the previous static skin model classifiers, dynamic classifiers, which are based on artificial neural networks (ANN) and/or genetic algorithms, have been proposed [16]. The flexibility and ability of ANNs to adapt to various image conditions make them a good choice for enhancing classification tasks for human skin pixels [7].

Most of the existing skin detection methods segment images using only skin color information. Based on the fact that the skin region of an image is a group(s) of homogenous connected pixels, texture information can be used to describe skin regions. Different texture descriptors, such as homogeneity, uniformity, and standard deviation, may be exploited for detection purposes [17].

Although color information plays the main role in modeling of skin, selecting the prober color space to present skin is crucial. Several comparisons between different color spaces used for skin detection can be found in the literature [18–22], but one important question still remains unanswered is, "what is the best color space for skin detection?" Many authors do not provide strict justification of their color space choice. Some of them cannot explain the contradicting results between their experiments or experiments reported by others [22]. Moreover, some authors think that selecting a specific color space is more related to personal taste rather than experimental evidences [18].

In this paper, a hybrid classifier that combines ANN with k-means clustering is proposed. The proposed classifier exploits the color and texture information to detect skin regions. This paper is organized as follows. In Section 2, an overview of related skin detection methods is presented. Section 3 describes the existing algorithms that form the background for the proposed method. The proposed skin detection method is described in Section 4. The experimental results are reported in Section 5; finally, a research discussion is presented in Section 6, and the conclusion is summarized in Section 7.

2. Related work

Artificial neural networks are interconnections of artificial neurons that, incredibly, mimic the organic neurons of the human brain. The main point of utilizing ANNs within skin color classification is to enhance the separability between skin and non-skin pixels. The first attempt to use an ANN for skin detection was made by Chen et al. [23]. They employed an ANN using the back propagation algorithm, utilizing the normalized RGB color space to reduce the sensitivity of illumination variations. Seow et al. introduced a skin color model based on an ANN used for face detection applications. They aimed to minimize the restriction relating to skin color variations between different races [24]. They were trying to create a self-adaptive color model based on a back propagation training algorithm and RGB color space. Yang et al. [25] employed a back propagation algorithm along with a Gaussian model classifier. They intended to enhance the accuracy of skin detection by using the Y component of YCbCr color space to overcome the impact of image illumination.

Zaidan et al. [26] proposed a module that hybridized a BP algorithm and heuristic rule method using YCbCr color information. The objective of their algorithm was to increase the classification reliability by removing the impact of illumination. The separability between skin and non-skin pixels relies on the output of the ANN and heuristic rules. Kakumanu et al. [27] adopted image chromatic adaptation. A neural network is used in this algorithm, which comprises two steps: first, to investigate the chromatic adaptation of the image, and second, to detect the human skin.

Duan et al. [28] applied the ability of Pulse Coupled NN (PCNN) to ease the imitation of the human vision technique by finding the relationships of the adjacent pixels. Bhoyar and Kakde [29] proposed a novel algorithm based on two output layer neurons, one each for the skin and non-skin classes. The novel method employed the Error Back Propagation Training Algorithm utilizing RGB color information. In this method, the threshold was used to overcome overlapping skin regions.

Doukim et al. [16] proposed numerous strategies using a Multi-Layer Perceptron ANN and the Y component from YCbCr color information. They employed a modified growing method to determine the number of neurons in the hidden layer. C-HN-O topology was employed in this work. A BP algorithm that combined the features of color and texture was proposed by Taqa and Jalab [17] to increase the reliability of the skin classification operation. Statistical models were used to estimate the texture features that increased the skin reliability.

The mentioned ANN-based techniques performed well in detection accuracy; however, these techniques have some drawbacks. All methods utilized a threshold value with no optimization process used to estimate these values, and the authors used the trial and error method, which is not an accurate procedure. In addition, of the mentioned algorithms, no real experiment was conducted using the ANN to select the color space. The selection relied on others' work. In addition to the drawbacks, the negative effect of illumination conditions and near skin color backgrounds represent a challenge that that must be overcome in order to develop reliable and robust detection algorithms.

k-Means clustering method used by Sree et al. to enhance the accuracy of skin detection [30]. In their algorithm, k-means clustering is used cluster the image into three clusters after using explicit rules. One of the three clusters contains skin regions and the other two contain the background and the edges of the skin regions. However, the authors did not explain how their algorithm can select the cluster that represents the skin area which is a major issue in their algorithm.

Another algorithm exploited k-means clustering for skin detection was proposed by Bevilacqua et al. [31]. In their algorithm, the image is first converted into CIE L*a*b color space, and then the image pixels are segmented into three clusters based on a and b channels. Supposing that one of the clusters will contain most of the skin pixels, the centroids of the three clusters are used to train an ANN. The objective of the ANN is to detect which cluster has the probability to contain skin. This neural network has six neurons at the input layer, an hidden layer of nine neurons and an output layer of three neurons. The output of the ANNis the probability that the cluster has skin. For example, if the output of the ANN is: (0.000020), (0.024), (0.98) it means that the third cluster has skin pixels. After that, the segmented image undergoes a cover-holes filter and a skin filter based on elliptical boundary models to enhance the output. The algorithm is based on the assumption that the most of skin pixels will be clustered in one cluster using a and b channels. However, this is not true for all images as the images vary in their background and illumination conditions. Fig. 1 shows an example of segmenting an image using k-means clustering based on ab channels. It is obvious that approximately half of the skin area is in cluster 1 and the other half is in cluster 2. That means half of the skin area will not be detected as the ANN will select only one cluster. As the result, the accuracy of detection will decrease dramatically.

3. Theoretical background

In this section, three methods, which form the basis for the proposed method, are described, namely: (1) MLP ANN, (2) Differential Evolution (DE), and (3) k-means clustering. In addition, this section presents the skin color and texture information that are employed in this paper.

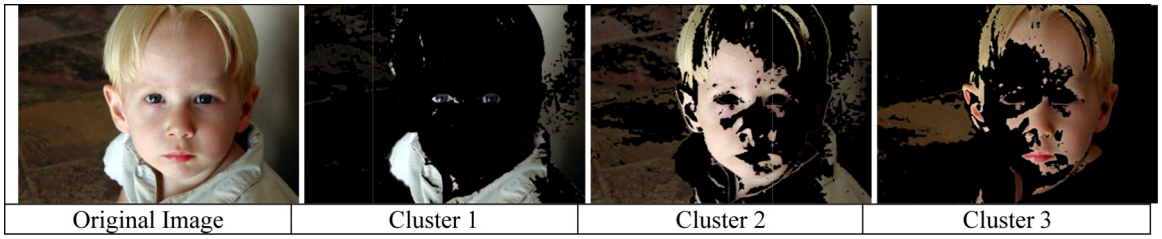

Fig. 1. Segmentation of image using k-means clustering based on ab channels ($k = 3$).

3.1. MLP ANN

In this work, the MLP was implemented using the MATLAB Neural Network Toolbox. For each configuration, the training process for each MLP structure was repeated 20 times with a different number of neurons to find the best performance (i.e., the global minimum) of the network error space in terms of minimum mean square error (MSE).

3.2. Differential Evolution (DE)

DE is a simple and fast heuristic search method that simulates the basic idea of organism evolution [32]. Using iterations, the DE method optimizes a given problem by trying to improve a candidate solution with regard to a given measure of quality. It is used for multidimensional real-valued functions without using the gradient of the problem being optimized. In contrast to classic optimization methods, such as gradient descent and quasi-Newton methods, DE does not require the optimization problem to be differentiable. Therefore, DE can also be used for optimization problems that are discontinuous, noisy and variable over time. In addition, the DE algorithm has less tunable parameters and has the ability to self-organize. The simplicity and ease of implementation make the DE algorithm very popular; this algorithm can be exploited in a wide range of applications such as digital filter design [33], shape reconstruction [34], and digital image watermarking [35].

When DE is used to optimize a function with n real parameters, the DE will generate a population of candidate solutions within predefined boundaries. All candidate solutions are tested and evaluated trying to locate the minima of the objective function. If the optimum solution is not reached, a genetic algorithm is used to generate another population. During each iteration, called a generation, new candidate solutions are generated by the combination of solutions randomly chosen from the current population (mutation). The outcoming solutions are then mixed with a predetermined target solution. This operation is called recombination and produces the trial solution. Finally, the trial solution is accepted for the next generation if and only if it yields a reduction in the value of the objective function, i.e. the stop criterion. This last operator is referred to as a selection. Fig. 2 illustrates the steps of DE method.

3.3. k-Means clustering method

k-Means is a popular unsupervised learning algorithm that is used in a wide range of applications, such as data mining, because of its simplicity [36]. The purpose of k-means clustering is to separate n observations into k clusters so that each observation is assigned to the cluster with the nearest mean. Each cluster has a centroid, and the centroids should be defined and distinct from each other. After defining the k centroids, the next step is to take each point belonging to a given data set and associate it with the nearest centroid. The groupage is completed when no point is pending. Then, k new centroids are recalculated as barycenters of the clusters resulting from the previous step. After that, a loop is used to bind the same

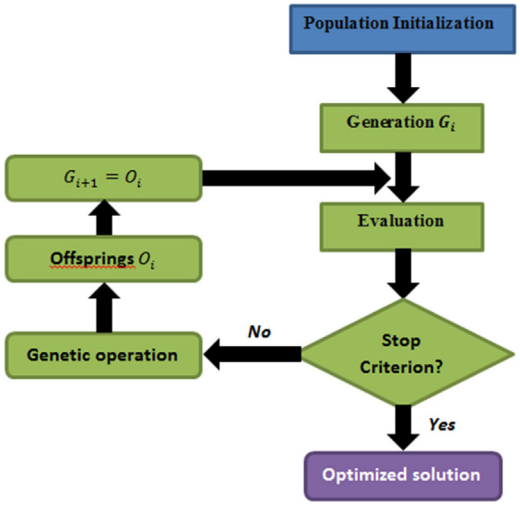

Fig. 2. The DE algorithm.

data set points and the nearest new centroid. During the loop, the k centroids change their location step by step, and the loop stops when no more changes can be made. In addition to its simplicity, k-means works well with large datasets. If k is small, k-means may be computationally faster than other techniques such as hierarchical clustering [37]. Moreover, k-means may produce tighter clusters than hierarchical clustering [38]. It can be employed in many fields, such as medical image segmentation [39], brain tumor detection [40], and content-based image retrieval [41].

3.4. Skin color and texture information

Digital image segmentation and classification are good examples of research areas that depend on color as a significant source of information. Nevertheless, some of the original colors are not suitable for certain types of image processing, and hence, adjustment of the color space is required for image analysis. For this adjustment, colors should be transferred from one space to another to obtain accurate results.

A skin detection process can be summarized in two steps: express the image using the perfect color space and classify the skin pixels based on the assigned skin samples using inference methods [2]. Selecting the proper color space is essential for skin color detection.

Different color spaces are compared for skin detection; comparison results show that there is no decision about the best color space used in skin detection. Most researchers do not provide good justification for selecting the appropriate color space [14] or for clarifying conflicts between the results of other authors [27]. Furthermore, some authors depend on their personal experience more than scientific experiments [28]. However, most researchers indicated that different skin detection techniques work differently using different color spaces [2,14]. In this work, six color spaces are investigated

using an MLP ANN to introduce the best color space used in skin detection. The color spaces used are RGB, YCbCr, YIQ, YDbDr, HSV, and CIE L*a*b.

4. The proposed method

This work proposes an optimized method for human skin detection. By optimized, we mean selecting the set of input features that can be optimally used by an MLP ANN to achieve the best detection accuracy. Developing the proposed method involves two phases: training an MLP ANN to detect optimized input features and enhancing skin detection using k-means and an MLP ANN. During the first phase, the DE algorithm is used to select optimal input variables as a combination of the color and texture information. In the second phase, the k-means clustering method is used to enhance the output of the optimized MLP ANN. The following sub-sections show the process of developing our method.

Based on the fact that the skin region of an image is a group(s) of homogenous connected pixels, texture information can be used to describe skin regions. In this paper, six different texture descriptors will be investigated: uniformity, standard deviation, skewness, kurtosis, smoothness, and entropy. The six texture descriptors were chosen for their simplicity and efficiency, which makes them suitable for real-time skin-based applications. Much more sophisticated texture descriptors may be exploited, such as Gabor filters, but they slow down the detection process. The formulas that can be used to calculate the six descriptors are given in Table 1.

4.1. Preparing the dataset

In preparation for training for an MLP ANN, three types of blocks (patches) are collected: pure skin blocks, pure non-skin blocks, and mixed blocks. Table 2 illustrates the details of the training data used for the training process.

Referring to Table 2, Humanae is a chromatic inventory that provides a dataset containing a broad collection of images of different people [43]. In addition to the diversity, the images of the Humanae dataset are of high resolution compared to the Compaq dataset. One hundred fifty images containing human skin were downloaded from the "Humanae Project" webpage, and then 40×40 pixels sized blocks were selected manually from different areas of the image, as shown in Fig. 3. The blocks were selected from the forehead, cheeks, shoulders, or chest. These locations were chosen to take into consideration any small differences in skin color tone that may exist between the different body areas of the same person.

The mixed blocks are blocks that contain both skin and non-skin pixels. They were selected from images of the ECU dataset in such a manner that they have almost an equal number of skin and non-skin pixels. The selection process was completed manually; achieving equal numbers of skin and non-skin pixels was difficult. However, the resultant data set contains 1,173,435 skin pixels out of 2,400,000 pixels (equivalent to 48.49% of the dataset).

The Humanae dataset was used to collect pure skin patches while the pure non-skin patches were gathered from self-collected images dataset. Although those two dataset have no ground truth, it is easy to generate the ground truth for the patched which is either all 1s or all 0s. However, the proposed algorithm extracted data from 4×4 sub-blocks blocks and in real situations the sub-blocks cannot be only pure skin or pure non-skin. So, to simulate real situation, a third type of patches which contains skin and non-skin pixels is needed. Those patches are collected from images of ECU dataset which already has ground truth images. The patches which have been collected cover wide range of human race and color and different illumination conditions.

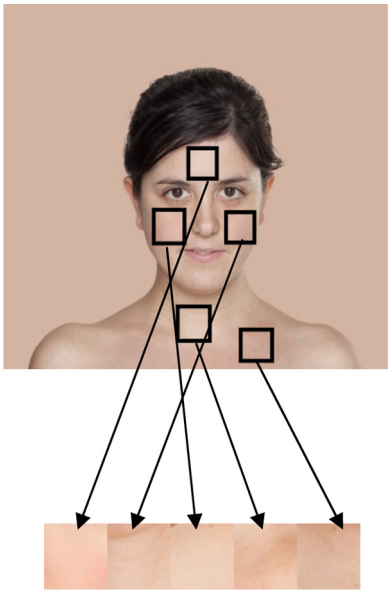

Fig. 3. Blocks extracted for training from Humanae dataset.

During the process of collecting the training data, each block is divided into sub-blocks of 4×4 pixels. For each sub-block, the RGB channels are separated and the corresponding 4×4 matrices are converted into three 16-element vectors. After that, the sub-block is converted into gray-level, and the six parameters; standard deviation, kurtosis, . . . etc., are calculated. For each sub-block, there will be six values that describes the texture of that sub-block. Since the proposed algorithm classifies pixels not blocks, the six values are repeated 16 times to form six 16-elements vectors. The color information, three 16-element vectors, and the texture information, six 16-elements vectors, are then concatenated into one 9×16 array. The process is repeated for all sub-blocks, and the resultant data is combined to form $9 \times 2{,}400{,}000$ array.

The process of collecting the training data from the blocks is illustrated in Fig. 4. It can be seen in the resultant matrix that R, G, and B values are different for each pixel, but the texture descriptors' values are the same. The process is repeated for each color space to obtain six different training datasets. In the same manner, for the mixed blocks, the corresponding ground truth of each block is divided into sub-blocks, and the binary values (1 for skin and 0 for non-skin) in the sub-blocks are converted into a 16-element vector. This vector will be the desired output during ANN training. The two processes are repeated for all sub-blocks and blocks, and the resultant vectors are concatenated in two arrays: input ($9 \times 2{,}400{,}000$) and output ($1 \times 2{,}400{,}000$).

4.2. Metrics

Skin detection in digital images can be considered as a classification problem, where image pixels, the dataset, are classified into skin pixels and non-skin pixels. Generally, the ratio of skin pixels to the total number of pixels in the images varies significantly. This makes the image an unbalanced dataset because the number of skin pixels does not equal the number of non-skin pixels. Measuring the detection accuracy as the total number of predictions that were correct may not be an adequate performance measure when the dataset is imbalanced [44]. Hence, we need appropriate metrics that reflect the actual accuracy. In this paper, we adopted precision (P), recall (R), and F_1-measure, which are commonly used measures in the field of information retrieval [45]. Precision is a measure of the probability that a detected region is correct, while recall is the

Table 1
Statistical texture descriptors used [42].

	Description	Formula	Notes
Standard deviation (σ)	A measure of average contrast.	$\sigma = \sum_{i=0}^{L-1}(z_i - m)^2 p(z_1)$	L is the number of possible intensity levels.
Uniformity (U)	The uniformity measure is maximum when all gray levels are equal (maximally uniform) and decreases from there.	$U = \sum_{i=0}^{L-1} p^2(z_i)$	z_i is a random variable indicating intensity.
Smoothness (r)	Measure of the relative smoothness of the intensity in a region. R is 0 for a region of constant intensity and approaches 1 for regions with large excursions in the values of its intensity levels.	$r = 1 - \frac{1}{1+\sigma^2}$	$p(z)$ is the histogram of intensity levels in a region.
Entropy (e)	A measure of randomness.	$e = -\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i)$	m is the mean (average) intensity.
Skewness (S)	A measure of the asymmetry of the intensity values around the mean (average) intensity.	$S = \frac{E(z_i-\mu)^3}{\sigma^3}$	$E(t)$ represents the expected value of the quantity t.
Kurtosis (K)	A measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3.	$K = \frac{E(z_i-\mu)^4}{\sigma^4}$	

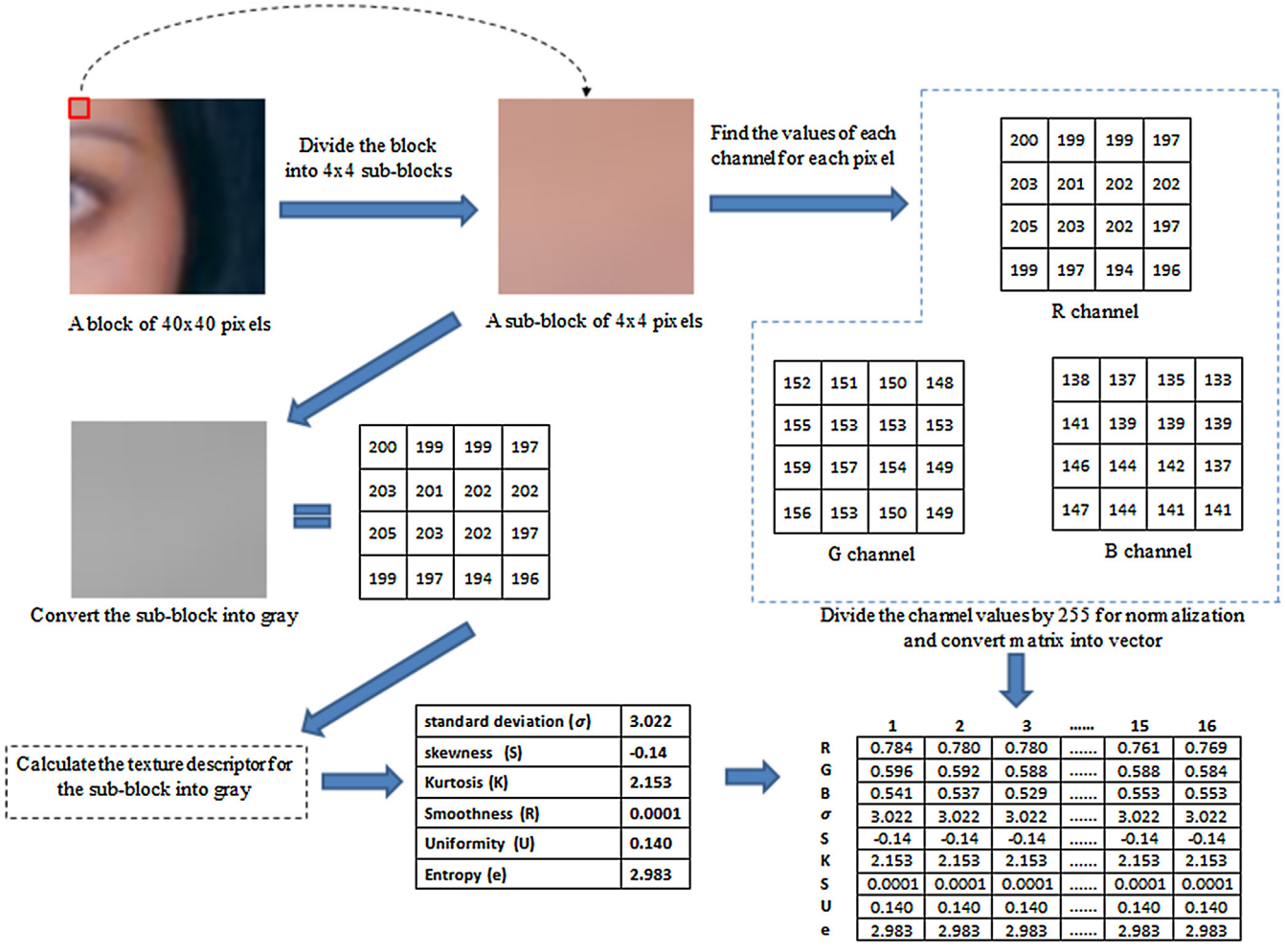

Fig. 4. Collecting training data from Humanae dataset blocks.

Table 2
Distribution of color-texture features in the training dataset.

	No. of blocks	Block size (pixels)	Total pixels	Source
Pure skin blocks	450	40 × 40	720,000	Humanae database
Pure non-skin blocks	450	40 × 40	720,000	Collected from the Internet
Mixed blocks	600	40 × 40	960,000	ECU database
Total size of database			2,400,000	

ratio of True Positive components to elements inherently ranked as the positive class. They can be calculated as follows:

$$P = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (1)$$

$$R = \frac{\text{true positive}}{\text{true positive} + \text{faslse negative}} \quad (2)$$

Because of the trade-off between P and R, the detection accuracy is measured using F_1-measure, which is the harmonic-mean of P and R. F_1-measure is given by:

$$F_1\text{-measure} = \frac{2PR}{P+R} \quad (3)$$

4.3. MLP ANN optimization

In order to exploit the DE algorithm for optimal selection, the chromosome and the fitness function should be designed appropriately. The chromosome's design and their evaluation to obtain our optimization goals are as follows:

(1) Optimization of objective function: There are two objectives in the optimization process: (a) to determine the best combination of color space, $C_1 \sim C_3$, and texture information, $T_1 \sim T_6$, that produces the highest detection accuracy and (b) to determine the optimal number of neurons n in the hidden layer of the MLP ANN.
(2) Chromosome design: We designed the chromosome so that all the parameters (i.e., n, $C_1 \sim C_3$, and $T_1 \sim T_6$) are simultaneously optimized. As shown in Fig. 5, the chromosome encapsulates the targeted parameters.

n	C_1	C_2	C_3	T_1	T_2	T_3	T_4	T_5	T_6
	R	G	B	σ	S	K	r	U	e

Fig. 5. The chromosome design consists of n, $C_1 \sim C_3$ (RGB color space), and $T_1 \sim T_6$ (texture descriptors).

During the optimization process, the color and texture parameters, $C_1 \sim C_3$ and $T_1 \sim T_6$, are assigned binary values 0 and 1. The binary value 1 indicates that the parameter is included within the current combination of color-texture training data and vice versa. The number of neurons in the hidden layer was investigated using 3–20 neurons. Based on Fig. 5, if the values of the current chromosome are 8 1 1 1 0 0 1 0 1 0, this indicates that the current MLP ANN has eight neurons in its hidden layer and five neurons in the input layer, and the training data are R, G, B, K, and U as shown in Fig. 6. For each training dataset, the training process is repeated for every chromosome in all generations as shown in Fig. 7. The maximum number of iterations used is 100, and the stop criterion is the minimum mean square error (MSE), which denotes the best performance of the MLP ANN.

After completion of the training process, the DE adjusts the combination to obtain a better quality meter (smaller MSE in our case) and performs another training process using the new input combination. The iterations continue until the DE algorithm finds the combination that yields the smallest MSE. By the end of the six optimization operations, the minimum MSE value was 4.46 when the YIQ color space was combined with standard deviation, kurtosis, uniformity, and entropy using 12 neurons. The results of optimization step is presented in Table 3.

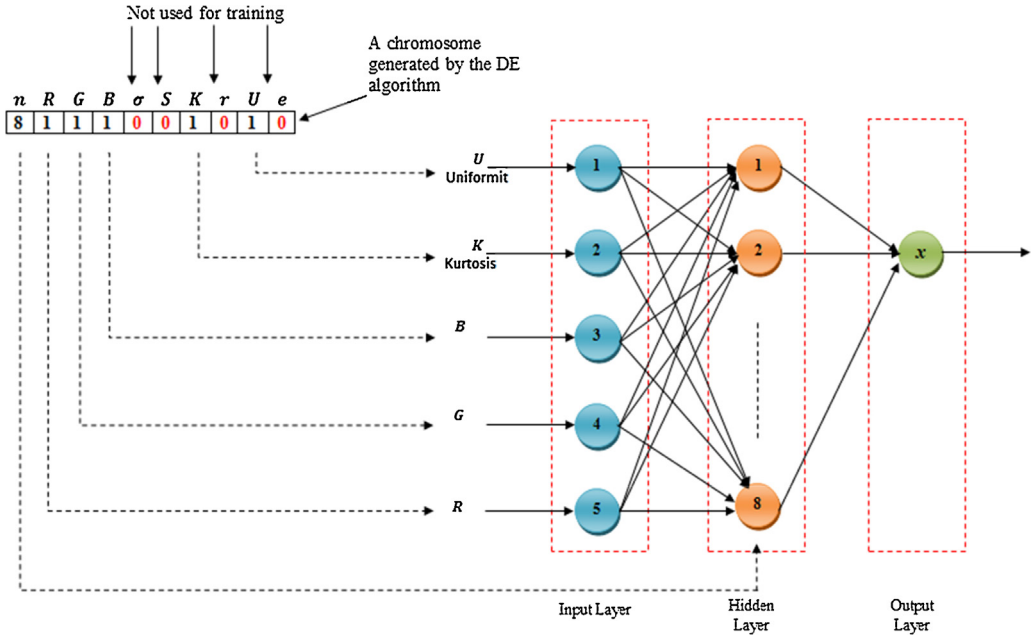

Fig. 6. ANN topology based on the chromosome generated by DE.

Table 3
Reults of optimization step: best color-texture combinations.

		Color spaces					
		RGB	YCbCr	**YIQ**	YDbDr	HSV	CIE L*a*b
Texture	σ	–	√	√	√	√	–
	S	–	√	–	√	√	√
	K	–	–	√	–	√	–
	r	√	√	–	–	–	√
	U	√	–	√	√	√	√
	e	√	√	√	–	√	√
Number of hidden neurons		8	8	**12**	10	12	12
MSE		4.78	4.66	**4.46**	4.57	4.51	4.61

Highlight the lowest values.

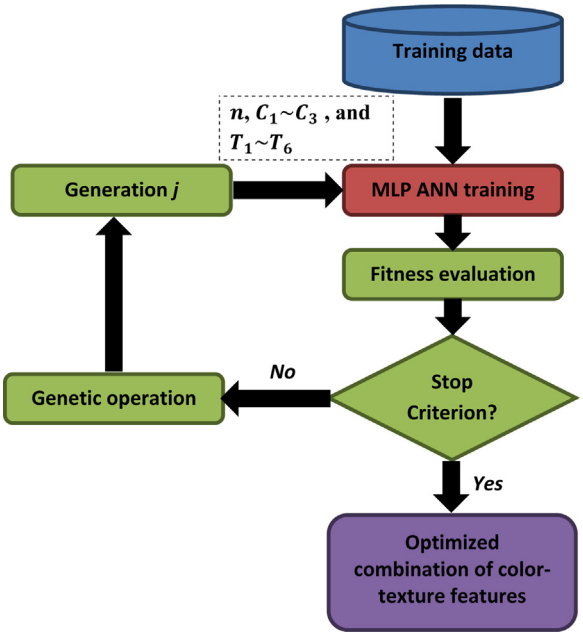

Fig. 7. Parameter optimization process.

4.4. Enhancing the output of the MLP ANN

From the output of the optimized MLP ANN, we can see that the image can be segmented into three areas: skin areas (white), non-skin areas (black or dark), and skin-like areas (between white and black) as shown in Fig. 8. This segmentation is because the output of the MLP ANN is not binary numbers but ranges from 0 to 1, where 0 denotes non-skin and 1 denotes skin. A common procedure is to use a threshold to convert the output of the MLP ANN to a binary image that presents the skin and non-skin region of the image. Instead of using a threshold, we opted to use the k-means clustering method to convert the output of the MLP ANN to the desired form.

The output of the MLP ANN is one-dimensional data where each pixel is represented by a single value; however, it will be better if that single value is paired with another value to describe or represent the pixel. With two-dimensional data vectors, segmenting the image into skin and non-skin by some clustering method, such as the k-means clustering method, can be much more accurate. To use the k-means clustering method, we investigated pairing the output of the MLP ANN with the information of one of the original image channels, such as *Y*, *Q*, *Cb*, and *H*. After testing different channels, we found that the *Q* channel of the YIQ color space gave the best results in terms of accuracy.

Clustering an image using the paired data along with the k-means clustering method into three clusters is shown in Fig. 8. The *x*-axis represents the output of the MLP ANN, and the *y*-axis represents the *Q* channel. For each cluster, the k-means method determines a centroid, which is the point with coordinates equal to the average values of the variables (output value and *Q* value) for the pixels in that cluster. Skin areas have a white color because the skin pixels have values of 1 or near 1. Hence, the cluster that represents skin is the one that has a centroid with the highest *x*-coordinates. Coordinates of the clusters' centroids for the image whose results are in Fig. 9 are listed in Table 3.

As shown in Table 4, the cluster that represents the skin area has the highest *x* coordinates (0.966). This fact can be used to identify the skin area after clustering pixels using the k-means clustering technique.

This is the advantage of combining the output image of the MLP ANN with the *Q* channel of the image. If the k-means clustering

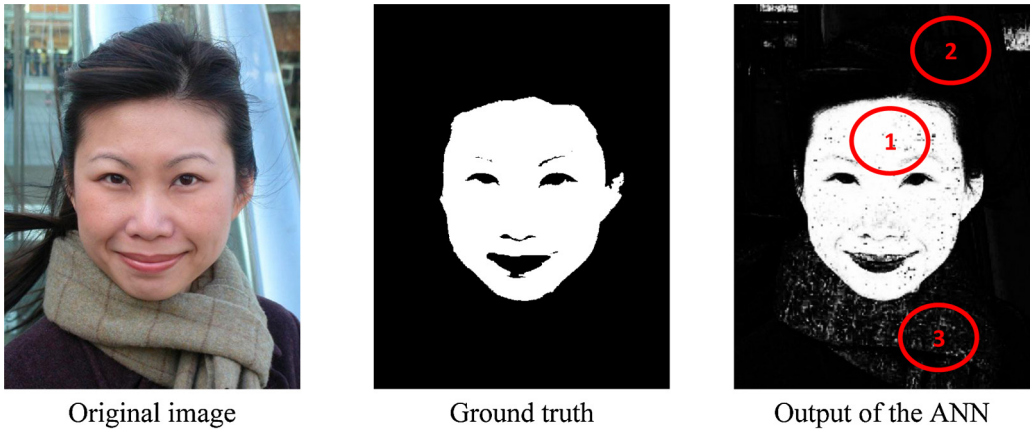

Fig. 8. Three areas in the output of the MLP ANN: (1) skin, (2) non-skin, (3) skin-like.

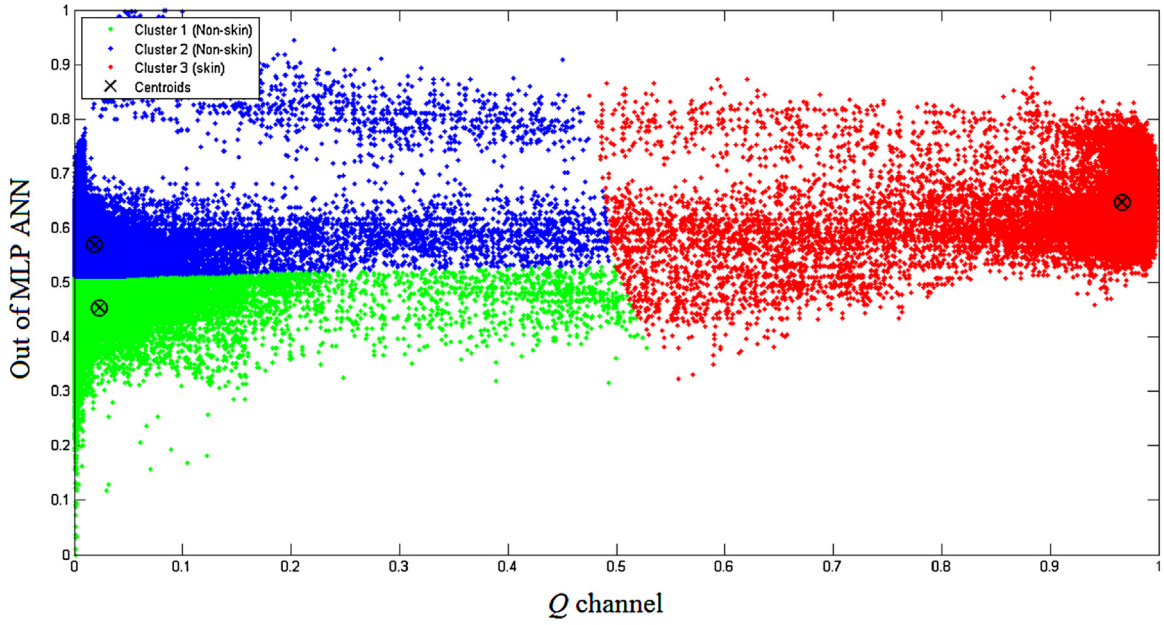

Fig. 9. Clustering the data into three clusters.

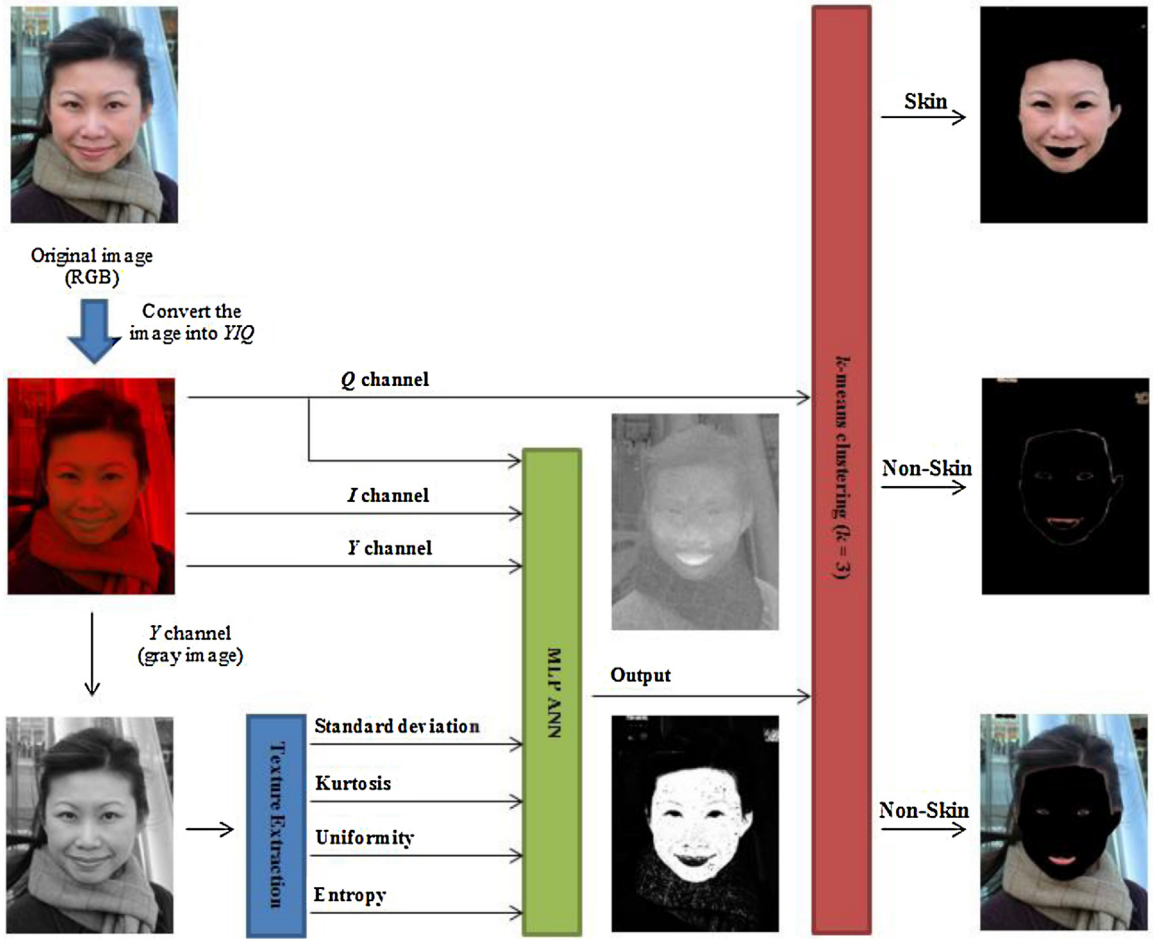

Fig. 10. Proposed skin detection method.

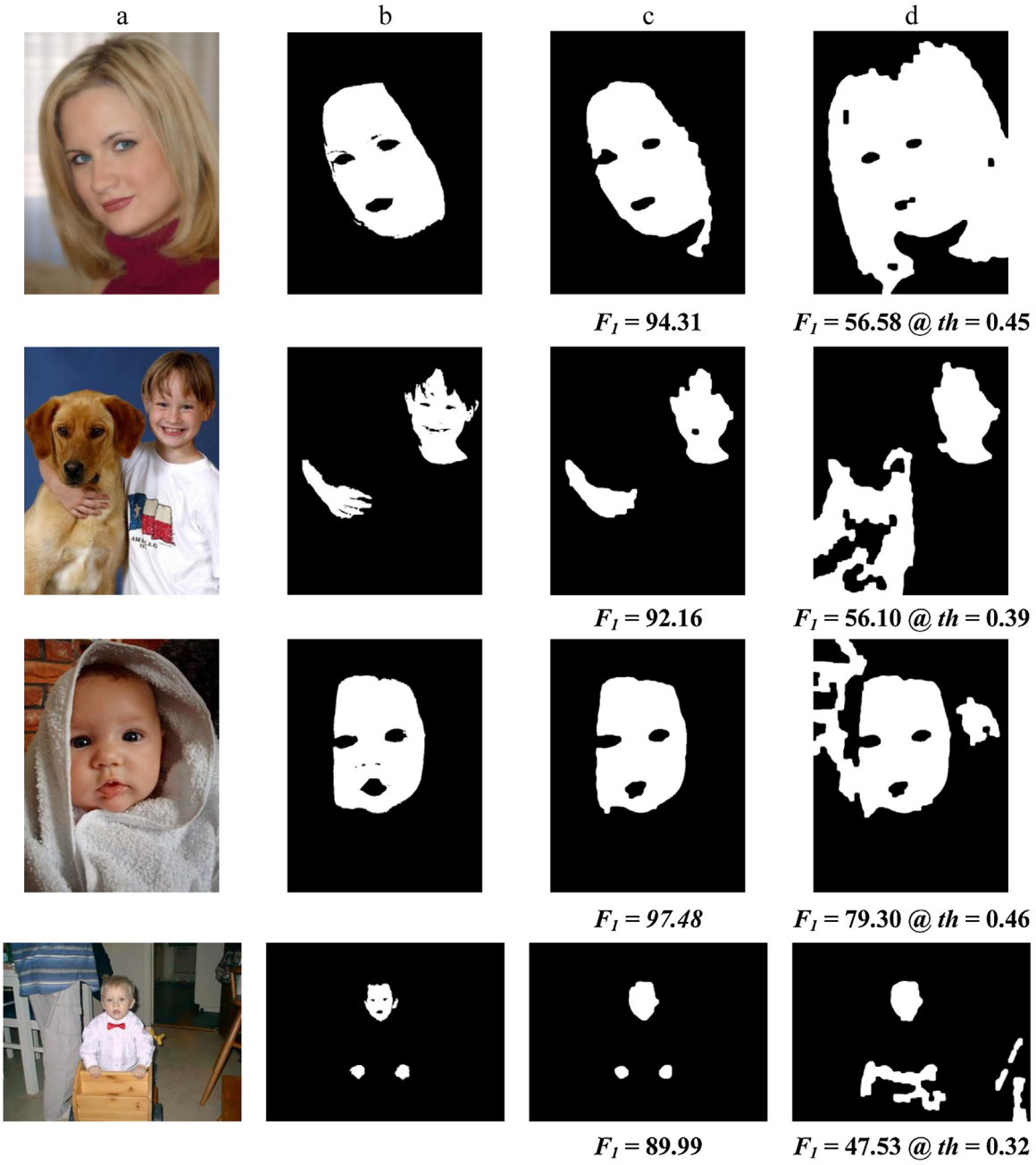

Fig. 11. Results of skin detection: (a) original image, (b) ground truth, (c) proposed detected skin (using MLP ANN with k-means), (d) detected skin (using MLP ANN with a dynamic threshold).

technique is used to segment the image into skin and non-skin using any two channels, such as *Y*–*Q* or *Y*–*I*, the cluster that corresponds to the skin area cannot be identified automatically. With the proposed algorithm, identifying the cluster corresponding to the skin area becomes direct and explicit.

The proposed algorithm can be summarized as follows:

Phase A: Image dataset training and optimized feature extraction

(1) Different color spaces and texture descriptors are studied and investigated.
(2) DE algorithm is used to select the best combination of color-texture information that can be used as input to the MLP ANN.

Phase B: Skin detection

(1) The image is converted into YIQ color spaces and then divided into 4×4 blocks.
(2) For each block, the texture descriptors (standard deviation, kurtosis, uniformity, and entropy) are extracted and combined with YIQ color information to form the input.
(3) The input data are applied to the optimized MLP ANN that was obtained in Phase A, and an output image *O* is obtained. *O* has the range (0–1).

Table 4
The three clusters and their centroids.

Cluster	Coordinates		Area type
	x (output of MLP ANN)	*y* (*Q* channel)	
1 (Blue)	0.019	0.569	Non-skin
2 (Green)	0.023	0.453	Non-skin
3 (Red)	0.966	0.647	Skin

Table 5
Results of skin detection using different methods.

Method	Recall (R, %)	Precession (P, %)	F_1 (%)
The proposed method (MLP ANN + k-means clustering)	88.00	87.65	**87.82**
MLP ANN + dynamic threshold [35]	85.51	79.32	82.30
Explicit rules (YCbCr + YUV) [9]	91.16	41.71	57.24
Explicit rules (RGB + YUV) [9]	89.62	59.12	71.25
ANN (RGB + texture) [17]	81.26	77.31	79.23
Bayesian network (YCbCr) [48]	62.87	41.28	49.83
Bayesian network (YIQ + texture)	85.79	50.82	63.83

Highlight the highest value.

(4) Then, O is combined with the Q channel that was obtained in step 1 to form the 2-dimensional data (O–Q).
(5) k-Means clustering technique is used to cluster OQ data into three clusters C, and each cluster has its centroid; the coordinates of the centroids Cx and Cy represent the O (x-axis) and Q (y-axis) correspondingly. The k-means technique labels the pixels in O according to the number of clusters to which they belong as follows:

$$O(C_i) = i, \quad i = 1, 2, 3 \tag{4}$$

(6) The number of the cluster corresponding to skin area, s, can be identified according to the following equation:

$$s = i \quad \text{if} \max(Cx) = Cx_i, \quad i = 1, 2, 3 \tag{5}$$

(7) Then, the non-skin area in the labeled image O is omitted using the following equation:

$$O(C_i) = \begin{cases} 1 \text{ if } i = s \\ \\ 0 \text{ if } i \neq s \end{cases}, \quad i = 1, 2, 3 \tag{6}$$

The resultant image is a binary image where the skin area corresponds to 1 values and the non-skin area corresponds to 0 values.
(8) Morphological operations (opening and closing) are used to overcome holes and the isolated pixels that are outliers (skin pixels mistakenly detected as non-skin and vice versa).
(9) Finally, skin areas with sizes less than 5% of the largest skin area in the image are neglected. Those areas are mostly incorrectly detected as skin and cannot be omitted using the morphological operations.

The proposed algorithm is illustrated in Fig. 10.

5. Experimental results

The experiments were carried out using images from the ECU database. ECU images were used because the images are acquired using uncontrolled lighting conditions, and objects with a skin-like color often appear in the background, which makes skin detection a difficult process [46]. The ECU images are provided with ground truth skin binary masks that indicated the skin regions.

The proposed algorithm achieved an accuracy of $F_1 = 87.82\%$ compared to the accuracy of $F_1 = 82.3\%$ that can be achieved when the MLP ANN is used along with a dynamic threshold as discussed in our previous work [47]. Fig. 11 illustrates a comparison between the proposed algorithm (MLP ANN with k-means clustering) and the MLP ANN with a dynamic threshold.

The proposed algorithm was compared with three different algorithms based on explicit rules [9], ANN along with RGB and texture information [17], and Bayesian network with YCbCr color space [48]. The comparison is presented in Table 4.

6. Discussion

From the table below, we observe that the proposed method has the highest accuracy, 87.82%, in terms of F_1-measure. The main contributor to that accuracy is the optimized MLP ANN used. Even when only a simple dynamic threshold was used, the MLP ANN achieved an accuracy of 82.30%, which is still higher than the other methods we used for comparison. The experimental results confirmed the superior performance of ANN over the other methods. Although the method in Ref. [17] is based on ANN and color-texture descriptors, it achieved only 79.23% because they used different color space and slightly different texture information (standard deviation, entropy, and range). This confirms the fact that color spaces have different abilities to represent human skin. In addition, color information only is not enough for accurate skin detection.

The high performance of the ANN-based methods, in terms of F_1-measure, is because of their ability to achieve a high and balanced recall and precision. In Table 5, the three ANN-based methods have recall and precision values of (88, 87.65), (85.51, 79.32), and (81.26, 77.31). On the contrary, the methods with explicit rules can achieve high recall values, 91.16 and 89.62, but low precision values, 41.71 and 59.12. Such unbalanced recall-precision values make the accuracy quite low. The same thing can be said about Bayesian Network-based methods.

The proposed method achieved very good performance; however, illumination conditions and objects or backgrounds with skin-like color are still a challenge. Poor illumination quality in dark images makes skin pixels look darker and result in misclassification as non-skin pixels, and this decreases the recall values. However, background pixels and objects with skin-like color are misclassified as skin, and this decreases resection values.

7. Conclusion

In this paper, a novel human skin detection method is proposed. The proposed method is a hybrid method that combines the advantages of two clustering techniques: neural networks and k-means. In addition, different combinations of color-texture descriptors were investigated to determine the optimal descriptor among the possible combinations. A powerful optimization method, DE, was used in order to determine the optimal combination that can be used for accurate skin detection. The experimental results show that the proposed method can achieve a high accuracy, with an F_1-measure of 87.82%, based on images from the ECU database.

For future work, we suggest employing a dynamic lighting correction algorithm in order to overcome poor illumination conditions. To overcome skin-like pixels, more texture descriptors may be investigated to find texture information that can represent human skin accurately.

Acknowledgment

The authors would like to acknowledge Universiti Sains Malaysia Research Grant Individual (USM-RUI) with No: 1001/PELECT/814092 for the financial support.

References

[1] C.C. Liu, P.C. Chung, Objects extraction algorithm of color image using adaptive forecasting filters created automatically, Int. J. Innov. Comput. Inf. Control 7 (10) (2011) 5771–5787.
[2] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection methods, Pattern Recognit. 40 (3 (March)) (2007) 1106–1122.
[3] C. Zhipeng, Face detection system based on skin color model, in: 2010 Int. Conf. Netw. Digit. Soc., May, 2010, pp. 664–667.
[4] Z. Zhang, H. Gunes, M. Piccardi, Head detection for video surveillance based on categorical hair and skin colour models, in: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 1137–1140.
[5] J.S. Lee, Y.M. Kuo, P.C. Chung, Detecting nakedness in color images, in: H.T. Sencar, S. Velastin, N. Nikolaidis, S. Lian (Eds.), Studies in Computational Intelligence, vol. 282, Springer, Berlin, Heidelberg, 2010, pp. 225–236.
[6] J. Han, G. Awad, A. Sutherland, Automatic skin segmentation and tracking in sign language recognition, IET Comput. Vis. 3 (1) (2009) 24.
[7] H.K. Al-Mohair, J. Mohamad-Saleh, S.A. Suandi, Human skin color detection: a review on neural network perspective, Int. J. Innov. Comput. Inf. Control 8 (12) (2012) 8115–8131.
[8] M. Abdullah-Al-Wadud, M. Shoyaib, O. Chae, A skin detection approach based on color distance map, EURASIP J. Adv. Signal Process. 2008 (1) (2008) 814283.
[9] F. Xiang, Fusion of multi color space for human skin region segmentation, Int. J. Inf. Electron. Eng. 3 (2) (2013) 172–174.
[10] H.K. Almohair, A.R. Ramli, A.M. Elsadig, S.J. Hashim, Skin detection in luminance images using threshold technique, Int. J. Comput. Internet Manag. 15 (1) (2007) 25–32.
[11] Son Lam Phung, Abdesselam Bouzerdoum, Douglas Chai, A novel skin color model in YCBCR color space and its application to human face detection IEEE International Conference on Image Processing, vol. 1, 2002.
[12] M.F. Hossain, M. Shamsi, M.R. Alsharif, R.A. Zoroofi, K. Yamashita, Automatic facial skin detection using Gaussian Mixture Model under varying illumination, Int. J. Innov. Comput. Inf. Control 8 (2) (2012) 1135–1144.
[13] B. Kwolek, Face tracking system based on color, stereovision and elliptical shape features, in: 2003 Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance, 2003, pp. 21–26.
[14] V.S.A. Vladimir Vezhnevets, Andreeva, A survey on pixel-based skin color detection techniques Proc. Graph., vol. 15, 2003, pp. 85–92.
[15] S. Khan, D. Bailey, G. Sen Gupta, S. Demidenko, Adaptive classifier for robust detection of signing articulators based on skin colour, in: 2011 Sixth IEEE International Symposium on Electronic Design, Test and Application, 2011, pp. 259–262.
[16] C.A. Doukim, J.A. Dargham, A. Chekima, S. Omatu, Combining neural networks for skin detection, Signal Image Process. Int. J. 1 (2) (2010) 1–11.
[17] A.Y. Taqa, H.A. Jalab, Increasing the reliability of skin detectors, Sci. Res. Essays 5 (17) (2010) 2480–2490.
[18] A. Abadpour, S. Kasaei, Comprehensive Evaluation of the Pixel-Based Skin Detection Approach for Pornography Filtering in the Internet Resources, 1996, pp. 1–6.
[19] B.D. Zarit, B.J. Super, F.K.H. Quek, Comparison of five color models in skin pixel classification, in: Proc. Int. Work. Recognition, Anal. Track. Faces Gestures Real-Time Syst. Conjunction with ICCV'99 (Cat. No. PR00378), 58–63.
[20] J.-C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, 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, 2000, pp. 54–61.
[21] G. Gomez, M. Morelos, On selecting colour components for skin detection, in: 16th International Conference on Pattern Recognition, 2002, pp. 961–964.
[22] D. Kuiaski, H.V. Neto, G. Borba, H. Gamba, A study of the effect of illumination conditions and color spaces on skin segmentation, in: XXII Brazilian Symposium on Computer Graphics and Image Processing, 2009, pp. 245–252.
[23] L. Chen, J. Zhou, Z. Liu, W. Chen, G. Xiong, A skin detector based on neural network, IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions, vol. 1, 2002, pp. 615–619.
[24] D. Valaparla, V.K. Asari, Neural network based skin color model for face detection, in: Proceedings of 32nd Appl. Imag. Pattern Recognit. Work, 2003, 2003, pp. 141–145.
[25] G. Yang, H. Li, L. Zhang, Y. Cao, Research on a skin color detection algorithm based on self-adaptive skin color model, in: 2010 International Conference on Communications and Intelligence Information Security, 2010, pp. 266–270.
[26] A.A. Zaidan, N.N. Ahmad, H.A. Karim, G.M. Alam, B.B. Zaidan, Increase reliability for skin detector using backprobgation neural network and heuristic rules based on YCbCr, Sci. Res. Essays 5 (19) (2010) 2931–2946.
[27] P. Kakumanu, S. Makrogiannis, R. Bryll, S. Panchanathan, N. Bourbakis, Image chromatic adaptation using ANNs for skin color adaptation, in: 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI, 2004, pp. 478–485.
[28] L. Duan, Z. Lin, J. Miao, Y. Qiao, A method of human skin region detection based on PCNN, Lect. Notes Comput. Sci. 5553 (2009) 486–493.
[29] K.K. Bhoyar, O.G. Kakde, Skin color detection model using neural networks and its performance evaluation, J. Comput. Sci. 6 (9) (2010) 955–960.
[30] P.K. Sree, I.R. Babu, Face detection from still and video images using unsupervised cellular automata with K means clustering algorithm, ICGST J. Graph. Vis. Image Process. 8 (2) (2008) 1–7.
[31] V. Bevilacqua, G. Filograno, G. Mastronardi, Face detection by means of skin detection, in: Fourth International Conference on Intelligent Computing, ICIC, 2008, pp. 1210–1220.
[32] P. Rocca, G. Oliveri, A. Massa, Differential evolution as applied to electromagnetics, IEEE Antennas Propag. Mag. 53 (February) (2011) 38–49.
[33] S.-T. Pan, B.-Y. Tsai, C.-S. Yang, Differential evolution algorithm on robust IIR filter design and implementation, in: Eighth International Conference on Intelligent Systems Design and Applications, 2008, pp. 537–542.
[34] I.T. Rekanos, Shape reconstruction of a perfectly conducting scatterer using differential evolution and particle swarm optimization, IEEE Trans. Geosci. Remote Sens. 46 (7 (July)) (2008) 1967–1974.
[35] V. Aslantas, An optimal robust digital image watermarking based on SVD using differential evolution algorithm, Opt. Commun. 282 (5 (March)) (2009) 769–777.
[36] K. Wagstaff, C. Cardie, Constrained K-means clustering with background knowledge, in: The Eighteenth International Conference on Machine Learning, 2001, pp. 577–584.
[37] A. Halder, Color image segmentation using rough set based K-means algorithm, Int. J. Comput. Appl. 57 (12) (2012) 32–38.
[38] D. Sonagara, S. Badheka, Comparison of basic clustering algorithms, Int. J. Comput. Sci. Mob. Comput. 3 (10) (2014) 58–61.
[39] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, W.L. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, vol. 2006, 2006, pp. 61–65.
[40] M.-N. Wu, C.-C. Lin, C.-C. Chang, Brain tumor detection using color-based K-means clustering segmentation, Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007) 2 (2007) 245–250.
[41] Z.S. Younus, D. Mohamad, T. Saba, M.H. Alkawaz, A. Rehman, M. Al-Rodhaan, A. Al-Dhelaan, Content-based image retrieval using PSO and k-means clustering algorithm, Arab. J. Geosci. (August) (2014) 1–14.
[42] R.C. Gonzalez, R.E. Woods, S.L. Eddins, Digital Image Processing Using Matlab, 2nd ed., Gatesmark Publishing, USA, 2009.
[43] Humanæ (2014) (Online). Available: http://humanae.tumblr.com/
[44] N.V. Chawla, Data mining for imbalanced datasets: an overview, in: O.Z. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Springer, US, 2005, pp. 853–867.
[45] C.D. Manning, P. Raghavan, H. Schutze, Introduction to Information Retrieval, Cambridge University Press, Cambridge, 2009.
[46] S.L. Phung, A. Bouzerdoum, D. Chai, Skin segmentation using color pixel classification: analysis and comparison, IEEE Trans. Pattern Anal. Mach. Intell. 27 (1 (January)) (2005) 148–154.
[47] H.K. Al-Mohair, J. Mohamad-Saleh, S.A. Suandi, Color space selection for human skin detection using color-texture features and neural networks, in: 2014 Int. Conf. Comput. Inf. Sci., June, 2014, pp. 1–6.
[48] A. Rasim, T. Alexander, Hand detection based on skin color segmentation and classification of image local features, TEM J. 2 (3) (2013) 150–155.