Skin Color Detection Using Stepwise Neural Network and Color Mapping Co-occurrence Matrix

Ghazali Osman, Muhammad SuzuriHitam
UniversitiTeknologi MARA, Malaysia
ghaza936@kelantan.uitm.edu.my
Universiti Malaysia Terengganu, Malaysia
suzuri@umt.edu.my

ABSTRACT
Skin color has been proven to be a useful and robust cue for face detection, human tracking, image content filtering, pornographic filtering, etc. Most of skin classification researches are focused on using pixel-based method to classify skin and non-skin pixels. This paper proposed a new technique for region-based skin color detection using texture information. The texture information was extracted from the color mapping co-occurrence matrix (CMCM). This technique is extension of gray level co-occurrence matrix (GLCM) which is introduced by Haralrick et. al to compute second order statistical texture features. The new color mapping matrix (CMM) between color bands have been developed for skin and non-skin area for each skin image and then, the CMCM were computed at four direction with distance, d = 1, and angle, θ = 0°, 45°, 90°, and 135°. The thirteen Haralick’s textures have been computed and used for formulating a skin color classifiers using stepwise neural network (SNN). The performance of each skin color classifier was measured based on true and false positive value. Besides that, the benchmark datasets from Universidad de Chile and TDSD were also be employed to test the skin color classifiers ability. The results shown that the skin color classifier formulated with [RGB] CMCM at direction (1, 0°) most superior as compared to other direction. Its average of true positive and false positive are 98.38 percent and 3.67 percent, respectively. Meanwhile, the classifier formulated with [RGB] CMCM at direction (1, 90°) is totally failed to classify skin and non-skin colors. Meaning that, the texture features which are computed from [RGB] CMCM at direction (1, 90°) cannot represent skin and non-skin color at all.

Indexing terms/Keywords
skin color detection; region-based classification; color mapping co-occurrence matrix; region-based skin color detection

Academic Discipline And Sub-Disciplines
Computer Science;

SUBJECT CLASSIFICATION
Image Processing; Artificial Intelligent

TYPE (METHOD/APPROACH)
Quasi-Experimental
INTRODUCTION

Skin is a largest organ of human body. Skin color is produced by combination of melanin, carotene, bilirubin, and hemoglobin. Different people have different skin color tone. The information of skin such as color and texture always used for clue for some application such as face detection [1], pornographic filtering [2], etc. Skin color detection has been used for pre-processing in these aforementioned applications.

The main challenge of skin color detection is to develop a skin color detection algorithm or classifier that is robust to the large variations in color appearance. Some color appearance variation such as skin appearance changes in color, intensity, and location of light sources, and other objects within the scene may cast shadows or reflect additional light. There are also many objects, which are easily confused with skin color. This is because skin-like materials are those that appear skin-colored under a certain illumination.

Skin color detection is a process to determine whether a desired pixel or a group of pixels belongs to skin or non-skin color. The presence of skin and non-skin can be determined by manipulating pixel color or pixels' texture. The objective of this paper is to develop skin color detection model based on skin region. The information of texture for each skin region which is computed from color mapping co-occurrence matrix has been used to formulate the model. The rest of the paper is organized as follows: Next section describes the existing skin color detection on region-based classification. METHOD section describes a proposed skin color detection method. Experimental results and discussion on region-based classification method proposed are presented in RESULT AND DISCUSSION section. Finally, main conclusions are outlined in CONCLUSION section.

BACKGROUND

From a classification point of view, skin color detection can be viewed as two classes problem, i.e. skin pixels and non-skin pixels classification. Pixel-based classification has long history. Starting from Jones and Rehg[3] till now, most of the researchers focused on this technique. Color and texture are two low-level features widely used for image classification, indexing, and retrieval. Color is usually represented as a histogram which is a first order statistical measure that captures global distribution of color in an image [4, 5]. One of the main drawbacks of the pixels-based approaches is that the spatial distribution and local variation in color are ignored. Another problem with pixel-based is false detection non-skin regions classify as skin due to similar color [6].

The Region-based classification technique is principally used to improve the results of the skin color modeling by reducing the false positive rate. This technique uses spatial arrangement information of skin color pixels to improve detection rates. The region-based classification takes into account the positional relative among pixels in a region of the image. It considers not only one pixel but also its behavior in an area that named texels (texture elements).

Region-based classification on texture feature is another important technique on image classification, which is different from color-based classification. Local spatial variation of pixel intensity is commonly used to capture texture information in an image. The texture features were extracted from region and applied to skin color detection model. In [7] proposed region-based algorithm for detecting skin color in static image. They had chosen the Single Gaussian Model skin color model in the normalized RG space after analyzing the distribution of skin color in six different 2D chrominance spaces, i.e. normalized RG, TS, HS, CIE-ab, IQ, and CbCr. Images are first segmented into paths using an improved fuzzy c-means algorithm and initialized cluster centroid, then, the percentages of skin color pixels in each patch can be obtained. According to corresponding percentage, patches are classified as skin color regions or not. They used a Receiver Operating Character (ROC) curve to analyze the performance of skin detection classifier with difference thresholds, and found that the true detection rate is 75.2 percent to 92.3 percent with false detection rate between 23.9 percent and 38.1 percent.

In [8] proposed skin detection algorithm using geometric features of skin regions to effectively classify skin and non-skin pixels. In this method, a non-parametric (histogram) skin color classifier is used for skin color detection. Then, the contours of skin regions are constructed using a curve evolution method based on adaptive grids. Finally, the geometric features are extracted from the contours and the cosine similarity measure is adopted for skin color detection. The experimental results have shown that the proposed method tested on faces, bikinis, and nude datasets performed well with skin detection rate of 88.6 percent to 93.5 percent and error rate of 3.75 percent to 6.2 percent.

In [9] proposed a skin color detection algorithm using a discrete Cosine transform (DCT). This algorithm is applied to each pixel location used a block-based feature vector which considers both color and texture information about skin's neighbors. The DCT coefficients are assumed to follow a generalized Gaussian distribution. The model parameters are estimated using the maximum-likelihood (ML) criterion applied to a set of training skin samples. The DCT of the neighborhood centered by each pixel is used to create the feature vector. For each pixel $p$, the DCT of a $3 \times 3$ block $M_p$ centered on $p$ is employed as the local region of pixels $p$. Subsequently, the DCT coefficients of each component block are computed separately, i.e. hue and saturation planes. To validate this technique, they used the Test Database for Skin Detection (TDS) [10] with manually labeled ground truth with true positive of 85 percent and 97 percent, and false positive of 10 percent and 30 percent, respectively.

In [11] proposed a Gabor filter along with a Sobel edge operator to improve skin color detection. A Gabor filter is band-pass filter that select a certain wavelength range round a centre wavelength using Gaussian function. They found that the performance of skin detection rate increase when the Sobel edged and Gabor filter applied as compared to pixel-based skin color classification. Meanwhile, in [12] used the Gabor filter to train skin and non-skin texture features. They found...
that many pixels whose color information is similar to the skin are removed because their texture features are different from those of the skin region.

Finally, in [13] compared some features for skin detection such as invariant moment, color histogram, statistic of the skin region, and texture. They applied energy, contrast, correlation, and entropy properties as texture features in their experiments. They found that texture based method performance better as compared to invariant moment, color histogram, and statistic of the skin region. Furthermore, they also proposed the combination of some weaker skin detection classifiers, which are built by the Learning Vector Quantization based on color, shape, texture features with Adaboost method. Adaboost is an adaptive algorithm to boost a sequence of classifiers. They also found that the combination skin color classifier perform well as compared to single skin color classifier.

**METHOD**

**Color Mapping Co-occurrence Matrix**

In [14] proposed gray level co-occurrence matrix (GLCM) to represent texture features in an image. In this method, the co-occurrence matrix is constructed based on the orientation, $\theta$ and distance, $d$ as illustrated by Figure 1.

![Figure 1: Geometry for measurement of gray level co-occurrence matrix for distances, $d$ with four angles, $\theta$](image1.png)

Similar to the concept of GLCM in computing texture features, this paper has proposed a new color mapping between the color bands of RGB. This is an extension of the method based on the GLCM proposed by [14]. This process will leads to four different matrices for skin portion and four different matrices for non-skin portion for each skin image as illustrated in Figure 2. This matrix is called color mapping matrix (CMM).

![Figure 2: Mapping directions between RGB color components](image2.png)
The new color mapping co-occurrence matrices (CMCMs) of each image were computed from these CMMs for skin and non-skin portion, separately. This color mapping were obtained for the distance of, \( d = 1 \) and angle, \( \theta \) from 0° to 135°. Thus, four directions have been tested for [RGB] CMCM, i.e. direction at (1, 0°), (1, 45°), (1, 90°), and (1, 145°).

The thirteen Haralick’s features (Table 2) were computed for each matrix. Each image will produce two sets of thirteen texture features whereby one set for skin color portion and another set is for non-skin color portion. Figure 3 illustrates the example of how the Haralick’s features compute from [RGB] CMCM matrix at (1, 0°).

![Figure 3: Illustration of the feature extraction from skin and non-skin pixels for one image](Note: R - red pixel; G – green pixel; B – blue pixel; CMCM – Color mapping co-occurrence matrix)

An example for CMCMs of skin and non-skin portions from Figure 3 shown by Table 1.

### Table 1: An example of skin and non-skin portion CMCMs

<table>
<thead>
<tr>
<th>CMCM skin color portion</th>
<th>CMCM non-skin color portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>R G B</td>
<td>R G B</td>
</tr>
<tr>
<td>: : :</td>
<td>: : :</td>
</tr>
<tr>
<td>: : :</td>
<td>: : :</td>
</tr>
<tr>
<td>95 14 6</td>
<td>1 5 9</td>
</tr>
<tr>
<td>95 14 6</td>
<td>2 6 10</td>
</tr>
<tr>
<td>95 14 6</td>
<td>3 7 11</td>
</tr>
<tr>
<td>95 14 6</td>
<td>4 8 12</td>
</tr>
</tbody>
</table>

(Note: R - red pixel; G – green pixel; B – blue pixel; CMCM- Color mapping co-occurrence matrix)

### Stepwise Procedure

The stepwise procedure is a common analytic procedure used in psychological and educational research to reduce the number of variables and to order variables in a given analysis [15]. It is most commonly employed in multiple regression and discriminant analysis. The goal of stepwise procedure is to sequence those variables (features) that maximise a
criterion, which describes their ability to separate classes from one to another while at the same time keeping the individual classes as tightly clustered as possible.

The variables criterion used for variables selection is Wilks’ Lambda, λ can be defined as follows:

\[
\Lambda = \frac{\det(W(z))}{\det(T(z))}
\]

where \(x_1, x_2, ..., x_2\) is a vector of the features that are currently included in the system. The \(W\) is the matrix of within- groups sum of squares and cross products for the features under consider is given by following equation:

\[
W(i,j) = \sum_{g=1}^{q} \sum_{t=1}^{n_g} (x_{igt} - \bar{x}_g)(x_{igt} - \bar{x}_g)
\]

(2)

The \(T\) is the matrix of total sum of squares and cross products is given by following equation:

\[
T(i,j) = \sum_{g=1}^{q} \sum_{t=1}^{n_g} (x_{igt} - \bar{x})(x_{igt} - \bar{x})
\]

(3)

where \(q\) is the number of classes, \(n_g\) is the number of samples in class \(g\), \(x_{igt}\) is the value of feature \(i\) for sample \(t\) of class \(g\), \(\bar{x}_g\) the value of feature \(i\) for sample \(t\) of class \(g\), \(\bar{x}_i\) is the mean of feature \(i\) over class \(g\), and \(\bar{x}_i\) is the mean of feature \(i\) over all classes, and \(\bar{x}_i\) is the mean of feature \(j\) over all classes.

### Back-Propagation Neural Network

A neural network (NN) is an information processing paradigm that is inspired by the way biological nervous systems such as the brain. It is composed of a large number of highly interconnected processing neurons working in unison to solve specific problems. A NN, like human being, it learns by example. Learning in biological systems involves adjustment to the synaptic connections that exist between the neurones. These neurones are connected each other by around \(10^{15}\) connections, creating a huge neural network [16]. Neurons send impulses make the brain works. The brain also receives impulses from the five senses and sends out impulses to muscles to achieve motion or speech.

There are three benefits of using NNs. Firstly, NN is trained by examples which mean no mathematical model of signals is to be estimated. Secondly, NN provides a non-parametrical method to approximate unknown systems, which can deal with not only statistical model but also non-linear models. This non-linearity is a very important property, which enhances the network’s classification or approximation capabilities without estimating any statistical parameter. Thirdly, their hierarchical and parallel structure also provides a speedy performance, which allows NN to be used in real time applications. The most critical advantage of using a NN is their adaptive learning capability, which enables NN to be taught to interpret possible variations of target objects. NN has the ability to learn complex data structure from a set of example patterns [17]. It has the advantages of working fast after the training phase even with large amount of data.

The back-propagation algorithm was introduced in [18], which adopted for estimating the dynamic models employed to predict nationalism and social communications. Generally, the scientific community remained unaware of the development of this important algorithm until the impressive research in [19] was published. Therefore, the back-propagation algorithms together with feed-forward networks were established as a major paradigm for use in the NN fields. In [20] introduced the concept of a recurrent-type NN which is considered by many as one of the more important contribution to both, the theory and application of a recurrent system. Neural network models based on a Hebbian learning [21] have become known as Hopfield nets. In [22] have suggested extensions to the concept of the Hopfield net, which is called the Boltzmann machine. Furthermore, in [23] were developed an adaptive bidirectional associative memory (BAM) network.

A number of simple highly interconnected signal processing units makes up a NN. These signal processing units are nonlinear mapping network that allows training and adaptively for a particular application. Classified by their interconnection architecture, NN can be classified in two types; multilayer perceptron (MLP) and recurrent NNs. The MLP is arranged in a feedforward manner in which neural nodes receive an input from external environment or other neural nodes, and pass the information to adjacent neural nodes without any feedback. Once MLP have been trained, the networks compute an output in response to the input pattern. Meanwhile, recurrent NN can be distinguished from the MLP in that they have a feedback loop in their networks structure.

Figure 4 illustrates the back-propagation NN architecture used for skin color detection in this paper. The values of input \((x)\) are textures features value and the output value \((y)\) is whether skin (1) or non-skin (0).
Figure 4: Neural network architecture for skin color detection

During the training stage, the pattern of performance and the mean square error (MSE) difference between the network target and actual output is computed to examine whether it has reached a criterion set. The number of an epochs and neural nodes in the hidden layer either should be increased or decreased to meet the MSE criterion. An epochs is defined as one complete of the entire training set during the learning process. If the MSE criterion is met then the training is completed; otherwise, the number of epochs will be increased. If increasing the number of epochs still cannot reach the MSE criterion, then it is necessary to increase the neural nodes in the hidden layer. However, increasing the number of hidden neural nodes results in a longer training period. The network was trained using Levenberg-Marquardt [24] back propagation algorithms. The performance of skin color classifiers formulated with CMC have been measured based on true and false indicators. The performance of classifiers also tests to the benchmarking datasets, i.e. TDSD and UChile. A stepwise procedure applied with NN, this method is called stepwise NN (SNN).

Skin Color Detection Performance Measurement

The performance of skin color detection has been measured based on true positive (TP) and false positive (FP) value. These value can be computed by following equation [25]:

$$TP = \frac{I_{pos}}{N_{pos}}, \quad FP = \frac{I_{neg}}{N_{neg}}$$

where, $I_{pos}$ is number of skin pixels of testing set correctly detected as skin, $N_{pos}$ is the total number of skin pixels in testing set, $I_{neg}$ is the number of non-skin pixels of the testing set falsely detected as skin, and $N_{neg}$ is the total number of non-skin pixels in testing set.

RESULT AND DISCUSSION

Stepwise procedure was applied to NN to choose a subset of texture features, which is sequentially, identify those textures that maximise a criterion to separate groups. The Wilks’ lambda criterion is used for this purpose. The output of NN is a non-linear equation which is used to classify skin and non-skin color.

Table 2 lists the texture features after stepwise procedure has been applied. Some of the textures which are not significantly contribute to classification were eliminated from the system.

Table 2: The list of significant texture features contribute to skin color classification

<table>
<thead>
<tr>
<th>Feature</th>
<th>Direction</th>
<th>$(1,0^\circ)$</th>
<th>$(1, 45^\circ)$</th>
<th>$(1, 90^\circ)$</th>
<th>$(1, 135^\circ)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Contrast</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Correlation</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Energy</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Entropy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Homogeneity</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Sum of square: Variance</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Sum Average</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Sum Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
9. Sum Entropy
10. Difference Variance
11. Difference Entropy
12. Information Measure of Correlation 1 (IMC1)
13. Information Measure of Correlation 2 (IMC2)

(Note: The (√) symbol indicates that texture features are selected after stepwise procedure)

Table 3 summarized the performance of Stepwise NN. The average of true and false positive rate was used to compare classifiers formulated between difference CMCM directions. Most of the results show that the performance of classifier formulated with [RGB] CMCM at direction (1, 0°) outperforms skin color detection rate. Meanwhile, the performance skin color classifiers formulated from [RGB] CMCM at directions (1, 45°) and (1, 135°) almost show the same. However, the classifier derived from orientation (1, 90°) is totally failed to detect skin and non-skin pixels.

Table 3: The performance of classifier formulated with [RGB] CMCM using stepwise NN

<table>
<thead>
<tr>
<th>CMCM at Direction</th>
<th>SlDb</th>
<th>UChile</th>
<th>TDSD</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
<td>FP</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>(1, 0°)</td>
<td>97.29</td>
<td>0.00</td>
<td>97.84</td>
<td>0.00</td>
</tr>
<tr>
<td>(1, 45°)</td>
<td>100</td>
<td>1.87</td>
<td>97.85</td>
<td>8.60</td>
</tr>
<tr>
<td>(1, 90°)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>(1, 135°)</td>
<td>97.20</td>
<td>1.87</td>
<td>97.85</td>
<td>5.38</td>
</tr>
</tbody>
</table>

(Note: TP – True positive; FP – False positive)

CONCLUSION
This paper discussed a region-based skin color classification technique using stepwise NN method to develop a skin color classifier. The region-based classification has been carried out using texture information extracted from color mapping co-occurrence matrix (CMCM) at four directions. It can be concluded that [RGB] CMCM at direction (1, 0°) is most superior as compared to direction (1, 45°) and (1, 135°). Meaning that the skin texture information that computed from CMCM at direction (1, 0°) strongly can be represented an information of skin and non-skin pixels portions. Meanwhile, the direction (1, 90°) totally cannot be used to model a skin color distribution because it provides a very low true positive rate. The technique proposed in computing a skin texture feature is providing a significant contribution to the performance of skin and non-skin classification.

ACKNOWLEDGMENTS
We wish to express my sincere appreciation to Professor Javier, Ruiz-del-Solar [26] from Universidad de Chile and Professor Hsu, Rein-Lien [27] from Michigan State University for their skin images database.

REFERENCES


AUTHORS

Ghazali Osman received his Diploma (Computer Science) from UniversitiTeknologi Malaysia, B.Sc. from UniversitiKebangsaan Malaysia, M.Sc. from UniversitiTeknologi MARA, Malaysia, and PhD from Universiti Malaysia Terengganu, Malaysia. He is currently a Senior Lecturer and the head of Study Centre, Faculty of Information Management, Faculty of Administrative Science & Policy Studies and Faculty of Laws, UniversitiTeknologi MARA, Kelantan Branch. His main research interests are in image processing, soft-computing, artificial intelligence, pattern recognition, and content-based information retrieval.
Muhammad SuzuriHitamobtained B.Tech (Hons.) Degree in Quality Control and Instrumentation Technology from UniversitiSains Malaysia and Ph.D from University of Leeds, U.K. He is currently an associate professor and the head of Computer Science Department, Faculty of Science and Technology, Universiti Malaysia Terengganu, Malaysia. His main research interests are in image processing, soft-computing and robotics.