



# An optimized skin texture model using gray-level co-occurrence matrix

Mahdi Maktabdar Oghaz<sup>1</sup> · Mohd Aizaini Maarof<sup>1</sup> · Mohd Foad Rohani<sup>1</sup> · Anazida Zainal<sup>1</sup> · Syed Zainudeen Mohd Shaid<sup>1</sup>

Received: 7 April 2017 / Accepted: 15 July 2017 / Published online: 28 July 2017  
© The Natural Computing Applications Forum 2017

**Abstract** Texture analysis is devised to address the weakness of color-based image segmentation models by considering the statistical and spatial relations among the group of neighbor pixels in the image instead of relying on color information of individual pixels solely. Due to decent performance of the gray-level co-occurrence matrix (GLCM) in texture analysis of natural objects, this study employs this technique to analyze the human skin texture characteristics. The main goal of this study is to investigate the impact of major GLCM parameters including *quantization level*, *displacement magnitudes*, *displacement direction* and *GLCM features* on skin segmentation and classification performance. Each of these parameters has been assessed and optimized using an exhaustive supervised search from a fairly large initial feature space. Three supervised classifiers including Random Forest, Support Vector Machine and Multilayer Perceptron have been employed to evaluate the performance of the feature space subsets. Evaluation results using Edith Cowan University (ECU) dataset showed that the proposed texture-assisted skin detection model outperformed pixelwise skin detection by significant margin. The proposed method generates an *F-score* of 91.98, which is satisfactory, considering the challenging scenario in ECU dataset. Comparison of the proposed texture-assisted skin detection model with some state-of-the-art skin detection models indicates high accuracy and *F-score* of the proposed model. The findings of this study can be used in various disciplines, such as face

recognition, skin disorder and lesion recognition, and nudity detection.

**Keywords** Gray-level co-occurrence matrix · Skin segmentation · Texture analysis · GLCM · Skin texture

## 1 Introduction

Proper detection of human skin is a vital preliminary task for many computer vision applications such as face recognition, skin disorder and lesion recognition, nudity detection and human tracking [1, 2]. Generally, skin detection methods utilize color information of individual or groups of pixels to discriminate the actual skin pixels from the non-skin ones. Although using the color information as the only discriminative feature leads to relatively fast skin detection, these methods are unable to efficiently discriminate the actual skin pixels in the presence of objects with similar color hue with human skin such as wood, animal fur and sand. To counter this issue, texture features are devised by considering the statistical and spatial relations among the group of neighboring pixels in the image instead of the color information of the individual pixels [3, 4]. Texture, which is defined as distribution of colors over a surface, delivers information about the spatial arrangement of the colors in an image region. Textures characterize the intensity or color variations that typically originate from roughness of the object surface. In general, texture analysis methods are categorized into three major classes including *structural methods*, *spectral methods* and *statistical methods*. Structural texture analysis methods describe the region's texture by using pattern primitives accompanied by certain placement rules. These methods are able to perform well in deterministic and predictable patterns

✉ Mahdi Maktabdar Oghaz  
Momahdi3@live.utm.my

<sup>1</sup> IASRG Lab, 4th Floor, Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor Bahru, Malaysia

(such as artificial man-made objects); however, the vast majority of the pattern in nature has some level of uncertainty and randomness and does not exhibit a constant behavior [3, 5–7].

Spectral texture analysis methods rely on autocorrelation of a region or power distribution of Fourier transformations to detect textural patterns. Fourier transformation is a robust technique in order to detect textures with periodic patterns; however, this technique is computationally expensive. Statistical texture analysis methods rely on textural features descriptors based on regional histograms and their moments. These methods have been used to measure the contrast, coarseness and granularity of a surface. Based on the order of the statistical function, texture feature descriptors are classified into two classes: first-order texture features and second-order texture features. The first-order texture features, known as Gray-Level Distribution Moments (GLDM), are only utilizing the intensity histogram information, with no attention to the spatial positioning of the pixels. Thus, this feature is unable to efficiently describe the textural patterns of a region. Unlike the first-order texture features, second-order texture features take into consideration the spatial positioning of the pixels relative to one another [8–10].

Gray-level co-occurrence matrix (GLCM) is one of the most prevalent statistical texture analysis methods which employ second-order texture features to analyze the regional textures. The GLCM characterizes the texture of an image by calculating how often pairs of pixels with specific values and in a specified spatial relationship co-occur in an image [5]. A co-occurrence matrix is a two-dimensional square matrix whose dimension corresponds to the number of gray levels in a grayscale image; for example, a 32-level grayscale image initiates a  $[32 \times 32]$  co-occurrence matrix. Each element  $(i, j)$  of the co-occurrence matrix indicates the number of times that pixels with value  $i$  and  $j$  are placed at the spatial position that the displacement vector refers to. Figure 1 demonstrates the construction of the GLCM given an eight-level grayscale

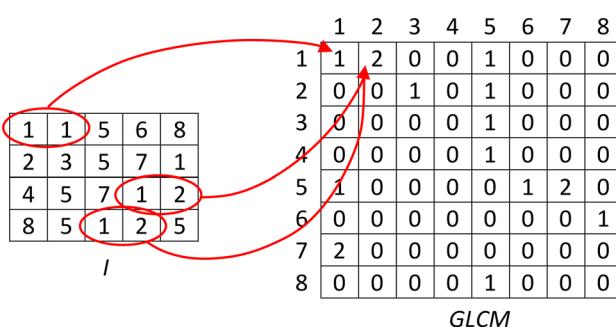
image  $I$  and displacement offset vector  $[0, 1]$  (immediate right neighbor). For example, element  $(1, 1)$  in the GLCM has a value of 1 because there is only one pair of ones in image  $I$ , placed at their immediate horizontal neighborhood. Element  $(1, 2)$  contains the value 2 since image  $I$  contains two pairs of horizontally adjacent pixels with values 1 and 2. In order to express the co-occurrence of the gray levels as probability distribution, the co-occurrence matrix should be normalized such that the sum of its elements is equal to unity. As a result, the GLCM will be independent of the size and dimension of the input image. To normalize the co-occurrence matrix, for each element  $(i, j)$  we have Eq. (1):

$$P(i, j) = \frac{C(i, j)}{\sum_{i,j=1}^N C(i, j)} \quad (1)$$

where  $C(i, j)$  is the value of the  $(i, j)$  element in the co-occurrence matrix,  $N$  is the number of rows and columns and  $P(i, j)$  is the normalized probability distribution of the element  $(i, j)$ .

There are several important parameters that affect the performance of the GLCM. The number of gray levels (quantization level) is one of the most important parameters in GLCM. More gray levels roughly provide more information about the textural feature of a region; however, this might increase the noise frequency and slows down the texture analysis process exponentially. The displacement magnitude of the co-occurring pixels is another important parameter in the GLCM. The choice of displacement magnitudes is very dependent on the surface texture. Coarse surfaces require relatively large displacement magnitudes, while fine surfaces demand smaller displacement magnitudes. The displacement direction of the co-occurring pixels is another key parameter in the GLCM [10]. The choice of displacement direction is very dependent on the orientation of the surface textures. Various displacement directions including 0, 45, 90 and 135 degrees can be used in order to disclose the presence of any vertical, horizontal or diagonal textural patterns. Several numeric features and statistics can be derived from the co-occurrence matrix. These features can be used to explicitly describe the textural properties of a surface. Entropy, energy, contrast, correlation and homogeneity are the most commonly used features of the co-occurrence matrix.

Despite the importance of the aforementioned parameters, numerous studies which have attempted to analyze human skin texture through GLCM features used arbitrary values to set up these parameters for skin texture analysis. Improper or arbitrary adoption of these parameters leads to ineffectiveness of the GLCM features in pixelwise skin detection [5, 6, 11]. To address this issue, this study investigates the effect of GLCM parameters, including



**Fig. 1** Image  $I$  and the corresponding GLCM with respect to  $[0, 1]$  displacement vector

quantization level, displacement vector magnitude and direction as well as GLCM features and statistics on skin texture detection and classification performance on photographic images. It is believed that by optimizing the aforementioned GLCM features and parameters, effectiveness of GLCM can be further improved for skin detection purposes.

The optimal value of each of these parameters was assessed using an exhaustive supervised search through a relatively large initial feature space (pool). Three supervised classifiers, Random Forest, Support Vector Machine and Multilayer Perceptron, were employed to evaluate the performance of the feature space subsets using an incremental exhaustive supervised search scheme. The exhaustive supervised search relies on Information Gain merit to incrementally populate the feature vector and returns the optimal feature space subset which denotes the proposed skin texture model. In order to evaluate the performance of the proposed skin texture model in the real world, the proposed texture model was integrated in a pixelwise skin detection as an auxiliary skin detection technique. The ECU (Edith Cowan University) dataset, which mimics the human skin real-world images, was used for evaluation and comparison purposes [12]. The performance of the texture-assisted pixelwise skin detection model was compared to the basic pixelwise model to identify how much improvement could be achieved using the proposed GLCM skin texture model. Comparison of the proposed model with state-of-the-art studies indicated satisfactory performance of the proposed model. The main contribution of this study is an optimized GLCM model for skin texture analysis, which is capable of improving the accuracy of pixelwise skin detection in photographic images. The proposed skin texture model can be achieved by optimizing GLCM parameters including quantization level, displacement magnitudes, displacement direction and GLCM features. The proposed model can be employed in wide variety of computer vision applications, such as face recognition, medical applications, nudity detection, and human tracking and surveillance purposes.

The remainder of this paper is organized as follows: Sect. 2 discusses the related work, Sect. 3 presents the proposed GLCM skin texture model, Sect. 4 presents evaluation results and analysis, Sect. 5 presents a comparison between the existing and the proposed skin texture models, and finally Sect. 6 summarizes the findings of this study.

## 2 Related work

The GLCM was initially introduced by [5] and has been extensively used for texture analysis purposes in various disciplines, such as Synthetic Aperture Radar (SAR)

[13, 14], material surface characterization [8, 15], biology, medical radiographic and ultrasound images and skin texture analysis in photographic images [16–23]. The use of GLCM showed a decent performance in texture analysis of human skin [18]. Numerous studies in this area are focused on detection of skin lesions and diseases.

Maurya et al. [18] used the GLCM in an automated skin cancer classification using a multiclass Support Vector Machine. They applied the autocorrelation, contrast, energy, entropy and homogeneity features of the co-occurrence matrix to construct the SVM feature vector. Pang et al. [23] also used GLCM texture features to analyze the effects of skin care products on human skin. Another *in vivo* study by Xiang Ou et al. [16] employed GLCM in order to describe and analyze the skin texture in skin capacitive images captured by a capacitance-based fingerprint sensor. In their study, angular second moment, entropy, contrast and correlation were the only GLCM features used to describe the skin texture. Another study by Gómez et al. [24] investigated the impact of different GLCM parameters such as distance, orientation, quantization level and GLCM features on the classification performance for breast ultrasound images, using Fisher linear discriminant analysis. De La Casa Almeida et al. [25] conducted a study on the efficiency of GLCM texture analysis in the assessment of skin cellulite using energy, entropy, contrast, homogeneity and textural correlation as the GLCM features. Another *in vivo* study by Zhang et al. [26] utilized the GLCM and Fourier transformation in order to differentiate the skin textures between young and old individual skin samples. Three texture features, energy, contrast and correlation, were obtained from the skin images and analyzed at four orientations (0°, 45°, 90°, and 135°). A study by Aswin et al. [27] paired GLCM features (contrast, correlation, energy, homogeneity) with color information (red, green, blue) in order to classify the skin cancer using an artificial neural network. In a similar fashion, Das et al. [28] used a GLCM alongside some auxiliary texture analysis techniques in order to identify skin diseases and lesions; however, unlike Aswin et al. [27], they employed a Support Vector Machine classifier to discriminate the skin texture.

The majority of these studies were carried out using skin textures images which were captured in controlled environments, in order to avoid the effects of image scale, orientation, lighting and environmental conditions, image quality and noise. Hence, their proposed skin texture models are very subjective and unable to properly analyze the skin texture in real-world uncontrolled scenarios. Apart from several applications of the GLCM in skin lesion and disease identification, this technique has been utilized in other disciplines such as SAR image analysis, biometrics, face recognition, pornography detection and hand gesture recognition.

Zhu et al. [19] used GLCM as auxiliary technique in pornographic image detection in order to decrease the disruption of the background region similar to skin area. Unfortunately, the GLCM parameters used in this study remain obscure. Pengyu and Jie [29] also used GLCM to detect pornographic images by means of key human parts. In a similar fashion, Wang et al. [30] used GLCM to identify the textural cues for pornographic image detection. Both of these studies assigned arbitrary values to GLCM parameters without providing salient reasoning. Yue et al. [31] used GLCM features for content-based image retrieval (CBIR) purposes, employing capacity, entropy, moment of inertia and relevance as the GLCM texture features. Despite this study not having a particular focus on skin texture analysis, their proposed technique can be applied to the skin texture field. Another study by Wang et al. [21] used GLCM for skin color detection in images with complex backgrounds, using contrast, angular second moment, entropy, correlation, homogeneity as the GLCM texture features. However, the authors used arbitrary values for other GLCM parameters such as quantization level, displacement vector magnitude and direction, without providing salient justification. A study by El-Khamy et al. [32] used GLCM features including energy, entropy and homogeneity for face recognition purposes, using the fast backpropagation neural networks model. Further details on other GLCM parameters remain obscure in this study. In a similar fashion, Al Abbadi et al. [33] employed GLCM features for skin texture recognition. Schwartz et al. [34] employed 12 GLCM features including angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy and directionality for human detection using Partial Least Squares Analysis. This study covers many GLCM features alongside a comprehensive performance analysis, although its main focus is on human detection rather than skin detection. A legacy study by Clausi and Jernigan [35] attempted to address the excessive computational burden of the GLCM through the implementation of a linked list algorithm rather than a traditional matrix-based approach. This study investigated the behavior of common co-occurrence texture features across different gray-level quantization levels. The authors claim that computational savings were substantial when using the linked list approach, relative to the matrix approach. In another study, Clausi [14] attempted to analyze the texture features of the SAR images using the linked list approach proposed by Clausi and Jernigan [35].

Despite the research efforts to initiate and adopt GLCM for skin detection and analysis, the majority of the existing studies contain only a limited number of GLCM features.

Furthermore, they have used arbitrary values to set up GLCM parameters such as quantization level, displacement magnitude and direction, with no solid justification. Thus, there is an evident lack of a comprehensive study which investigates the relevance and reliability of the GLCM features and parameters in uncontrolled photographic skin images.

Apart from the texture analysis techniques, various classification strategies have been employed for skin segmentation purposes. The majority of these studies used supervised classification techniques such as SVM [27, 36–38], decision trees [38–41], Naïve Bayes [36, 42, 43] and MLP [36, 38, 44–46] for skin segmentation purposes. Despite their differences, these techniques can achieve relatively comparable and somewhat satisfactory performance. With advancements in processing speed, computational resources and the development of parallel and GPU computing, novel machine learning techniques such as Deep Learning have emerged to address sophisticated classification issues such as image segmentation and object recognition. In recent years, Deep Learning techniques have also been utilized in various skin and face recognition applications, such as skin disorders and lesion detection.

Romero-Lopez et al. [47] proposed skin lesion classification from dermoscopic images using Deep Learning techniques. Their proposed solution is built around the VGG (Visual Geometry Group) convolutional neural network architecture and uses the transfer learning paradigm. In another study, Esteva et al. [48] used single deep convolutional neural networks and end-to-end training using pixel information for skin cancer classification. In similar studies, Liao [49], Jafari et al. [50] and Li et al. [51] employed the deep convolutional neural networks (CNN) approach for skin disease and Melanoma classification. Liu et al. [52] used Deep Learning techniques for identification of face attributes in photographic images. They employed two cascade CNN, where the first CNN was used for face localization and the second one was used for face attribute identification. Apart from the medical applications of Deep Learning in skin detection and segmentation, some other researchers, such as Bharati et al. [53], have used this technique for detecting facial retouching, age estimation [54] and pornography video detection [55]. A study by Zhang et al. [56] used a deep convolutional neural network to identify human gender in video streams, using facial features, with the aim to embed content-relevant ads within a given video. Another study by Zhang et al. [57] employed single recurrent neural networks for translating English subtitle characters to Chinese. Their proposed approach can be adopted for skin and face detection applications.

### 3 Proposed GLCM skin texture model

As mentioned earlier, this study conducted a comprehensive analysis of skin texture in photographic images by optimizing GLCM parameters including quantization level, displacement magnitudes, displacement direction and GLCM features. The efficiency and relevance of these four parameters were assessed through an exhaustive supervised feature selection technique. Through this study, we aimed to find the answers to the following questions:

- What is the optimal displacement magnitude for skin texture segmentation?
- What is the optimal displacement direction for skin texture segmentation?
- What is the finest amount of quantization level for skin texture segmentation?
- What are the most prominent GLCM features for skin texture segmentation?
- Do GLCM texture features improve skin segmentation performance?

#### 3.1 Dataset setup

The experiments in this study were carried out using two distinct datasets. The first dataset was an in-house two-class dataset, consisting of 1000 texture patches, 500 of which represented human skin texture. These skin texture patches were manually cropped from photographic images of over 100 different individuals. These images were taken by our research team using a DSLR camera to maintain fine skin texture patterns. Manual crop also ensures the reliability and trustworthiness of skin texture patches. Skin patches were evenly selected and cropped from different ethnic groups and skin color tones, to avoid bias toward any ethnic groups or skin color. The skin texture patches in this dataset were chosen from the various body parts, such as face, leg, trunk and arms, so that each body part has a relatively equal contribution to the final skin texture model. In order to better mimic the real-world scenario, the skin patches come in various scales, directions and lighting conditions. The remaining 500 texture patches which were collected and manually cropped from high quality Web images represented the texture of wooden surfaces, sand and animal furs which had high degree of similarity in color hue with actual skin regions. This allowed us to evaluate the effectiveness and performance of the GLCM texture features in challenging scenarios. All the texture patches were captured in camera raw format and then converted to lossless Tagged Image File Format (TIFF) using Adobe Light Room, to avoid any alteration in natural texture by compression algorithms such as JPEG. For the

same reason, all types of preprocessing and image enhancement were avoided. Patches were manually cropped to  $150 \times 150$  dimension, so each of them had an equal amount of contribution in the model. This dataset was mainly used to identify the optimum values of GLCM parameters and the feature selection process. Figure 2 illustrates some samples of the skin patches dataset.

The second dataset used in this study, named ECU (Edith Cowan University) [12], is a benchmark dataset in skin deletion consisting of 4000 images exposing human face and skin regions. The images from this dataset are associated with ground truth skin binary masks, which indicate the true skin regions; as a result, this dataset can be used to evaluate and compare the performance of the proposed skin texture model. ECU images were acquired in uncontrolled lighting conditions, and skin-like color objects usually appear in the image background scene, which challenges the skin segmentation. Figure 2 shows some sample images of the ECU dataset.

#### 3.2 Displacement magnitude and direction

The optimal value for displacement magnitude and direction is very dependent on the surface texture. Fine textures demand relatively small displacement vectors, while only larger displacement vectors are capable of revealing the texture properties of a coarse surface; however, this argument can only apply to images taken under controlled environment. Since human skin has a relatively soft and monotonous texture, it is expected that smaller displacement magnitudes are enough to reflect the skin texture. Unfortunately, there is no established technique to estimate the right amount of displacement magnitude and direction. A visual inspection of the behavior of some GLCM features across different displacement magnitudes and directions gives us a rough estimation of a suitable range for displacement magnitude and direction value for skin texture segmentation. Figure 3 illustrates the temperature diagram of Euclidian similarity of GLCM features in a window of  $[20 \times 20]$  pixels centered to the reference pixel, i.e., it shows the variations and behaviors of GLCM features in different directions and magnitudes of 10 pixels from the reference pixel. Skin texture patches from our in-house dataset were used to carry out this experiment.

The Euclidian similarity was assessed through Eq. (2) where  $Ed(R, O)$  is the Euclidian similarity between reference pixel  $R$  and displacement pixel  $O$ .

$$Ed(R, O) = \sqrt{(R - O)^2} \quad (2)$$

From Fig. 3, it can be inferred that GLCM features in skin texture are more likely to be a function of

Skin Patches Dataset Samples



ECU Dataset Samples

**Fig. 2** Skin patches and ECU datasets samples

displacement magnitude rather than displacement direction. GLCM features have analogous behavior across different directions in a relatively consistent manner without noticeable stress in a certain direction. In other words, we can state that displacement direction does not have a significant effect on skin texture segmentation using GLCM features in uncontrolled photographic images. As a result, this study only investigates the skin texture in direction  $\theta = 0^\circ$ . Apart from displacement direction, Fig. 3 also implies some useful information about displacement magnitude. From Fig. 3, it can be observed that depending on the GLCM feature, the reference pixel has relatively high degree of similarity/contrast with its immediate adjacent pixels. This similarity/contrast will gradually fade away as displacement magnitude increases to roughly 7 pixels. For displacement magnitudes above 7 pixels, the degree of similarity/contrast between the reference and displacement pixels is relatively negligible. Accordingly, this study investigated the skin texture in the displacement magnitude range of 1–7 pixels. It is anticipated that the candidate displacement magnitudes and direction are adequate to discover the skin texture in uncontrolled photographic images.

### 3.3 Quantization level

Quantization (number of gray levels) is another essential parameter in texture analysis using GLCM. This study examines human skin texture at various quantization levels, including 8, 16, 32, 64, 128 and 256 gray levels. Fewer gray levels fade the soft and monotonous textures away by discretizing the grayscale domain into bulkier bins. While more gray levels might describe the soft textures better, this increases the noise patterns as well as computational cost. Equation (3) shows how we quantized the images into different numbers of gray levels.

$$\text{int}_n = n * \text{round}(\text{int}_{256}/n) \quad (3)$$

where  $n$  is the number of gray levels in the quantized image,  $\text{int}_{256}$  is the input 256 gray-level image and  $\text{int}_n$  is the output image with desired number of gray levels. Figure 4 shows an example of 8, 16, 32 and 64 gray-level images from the ECU dataset.

### 3.4 GLCM features

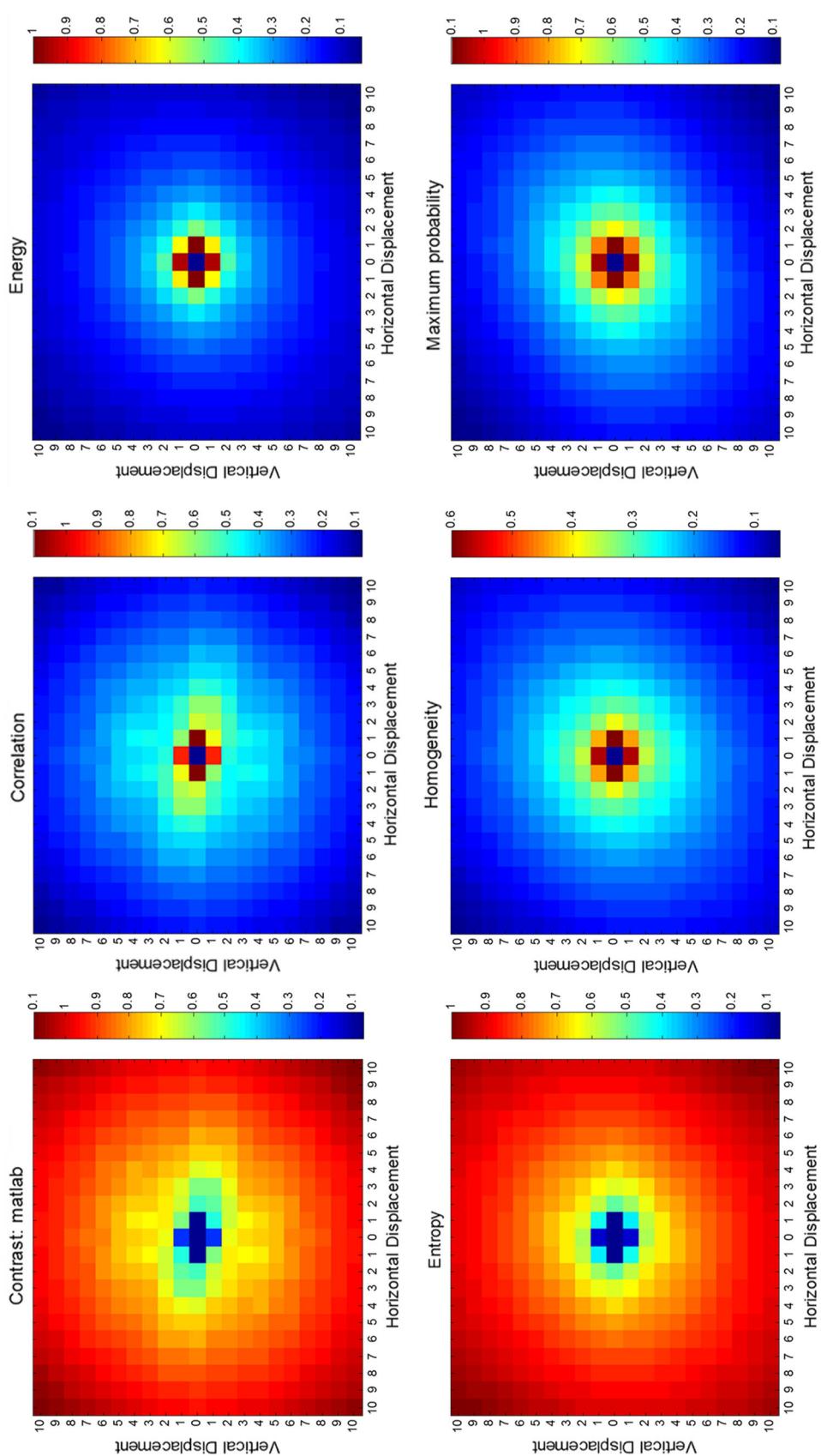
This study has employed 20 commonly used GLCM features found in the literature in order to assess the textural features of the human skin. Some of these features, like contrast and homogeneity, disclose the presence of a specific organized textural pattern within the image, while other features characterize the complexity and nature of gray-level transitions which occur in the image. These features, alongside a brief description, are listed in Table 1. The feature numbers in the first column have been used to address the corresponding feature in the rest of this manuscript.

## 4 Evaluation results and analysis

### 4.1 Analysis of GLCM parameters

The feature space in this study consists of total number of 140 features (7 displacement magnitudes (1, 2, 3,...,7)\*, 1 displacement direction ( $\theta = 0^\circ$ )\* and 20 GLCM features = 140) which have been examined at six different gray levels including 8, 16, 32, 64, 128, 256. Considering the likelihood of many noisy, redundant, correlated and irrelevant features in this feature space and the excessive amount of time required to deal with such a large feature space, the urge for a robust feature selection process is tangible. Correspondingly, this study employs a feature selection technique which performs an exhaustive search through all feature space subsets and evaluates each subset's feature by a supervised machine learning technique to

**Fig. 3** Temperature diagram of Euclidian similarity of GLCM features, including contrast, correlation, energy, entropy, homogeneity and maximum probability, in window of  $[20 \times 20]$  pixels centered to the reference pixel





**Fig. 4** Example of 8, 16, 32 and 64 gray-level images from ECU dataset

filter out non-informative and irrelevant features from the feature space. Exhaustive search might not be the fastest technique to find the dimensionally reduced optimal feature space, although it always promises to find the most accurate results. This study used three supervised classifiers, Random Forest, Multilayer Perceptron and Support Vector Machine, to evaluate the performance of each feature subset. Random Forest classifier has relatively high generalization capability and a fast training time. The number of decision trees in the Random Forest classifier was empirically set to 17 trees. Other parameters for Random Forest classifier are similar to the parameters in Khan et al. [38]. SVM classifier benefits from polynomial equations up to exponent three to build its kernel. The tolerance parameter was set to 0.001, epsilon for round-off error was set to  $10^{-12}$ , and the complexity parameter was set to 1. To set up the MLP classifier, a network of seven layers was used, which consists of the input layer, which takes the input data from GLCM feature space, five hidden layers and the output layer, which indicates the skin and non-skin classes. Neuron counts in the hidden layers were measured by averaging the input and output feature counts.

To validate the performance of each feature subset, tenfold stratified cross-validation technique was used for all classifiers. The dataset of texture patches was used for training and evaluation purposes. Performance metrics including *F*-score, accuracy, TPR and FPR were measured for both skin and non-skin classes and the average values reported as the figure of merit in this experiment. Throughout this manuscript, the features are addressed in the XXX\_YY\_ZZ scheme, where XXX indicates the quantization level, YY stands for displacement magnitude and ZZ represents the GLCM feature number in Table 1 (for example, “032\_07\_14” means sum entropy of 32 gray levels with displacement magnitude of 7 pixels) to simplify and shorten addressing of the features. In order to reduce the elapsed time required to find the optimal feature space subset through the exhaustive supervised search method, we devised a preliminary feature evaluation technique

which allows us to get rid of some non-informative, noisy and irrelevant features. This preliminary feature evaluation dramatically reduces the elapsed time to find the optimal feature space subset for skin texture evaluation. To do so, this study devised a feature ranking technique which evaluates each individual feature by measuring the amount of information gain with respect to the corresponding class. Information gain tells us how important a given feature of the feature space is. Gain is the difference between the entropies (unpredictability) of classes and features with respect to each class [58–60]. Equations (4) and (5) show how the information gain merit was measured for each GLCM feature.

$$\text{Ent}(T) = \sum_{i=1}^C -P_i * \log_2 P_i \quad (4)$$

$$\text{Gain}(T, F) = \text{Ent}(T) - \sum_{\partial \in \text{values}_F} \frac{|T_\partial|}{|T|} \text{Ent}(T_\partial) \quad (5)$$

where  $C$  denotes the number of different classes in the dataset,  $P_i$  is the proportion of S belonging to class  $i$ ,  $\text{Ent}(T)$  represents the Entropy of the given data, value  $F$  is the set of all possible values for feature  $F$  and  $T_\partial$  is the subset of  $T$  for which attribute  $F$  has value  $\partial$ .

Based on above definition, information gain merit was measured for each feature, which roughly indicates the discriminatory power of that feature for skin texture classification. Figure 5 illustrates the information gain merit of all GLCM features listed in Table 1, in ascending order.

Figure 5 shows that sum average, autocorrelation, sum variance, variance, entropy, sum entropy, cluster shade, energy and maximum probability have the highest information gain merit which makes them the most prominent GLCM features for skin texture classification. On the other hand, features like correlation, inverse difference and homogeneity have the least information gain merit and are prone to increase the data noise. An incremental exhaustive supervised search technique has been devised to find the optimal feature vector (feature selection process). This

**Table 1** GLCM features alongside the corresponding mathematical equations and brief description

No.	Feature name	Equation	Description
1	Autocorrelation	$\sum_{i,j=1}^N (ij)P(i,j)$	Returns probability occurrence of the specific a pixel
2	Contrast (inertia)	$\sum_{i,j=1}^N P(i,j)(i-j)^2$	Returns the local variation boundary of the gray levels in a texture
3	Correlation	$\sum_{i,j=1}^N (i-\mu_i)(j-\mu_j)P(i,j)/\sigma_i\sigma_j$	Returns the probability occurrence of the specific pixel pair
4	Cluster prominence	$\sum_{i,j=1}^N (i+j-\mu_x-\mu_y)^4 P(i,j)$	Measures the skewness of the matrix (asymmetry)
5	Cluster shade	$\sum_{i,j=1}^N (i+j-\mu_x-\mu_y)^3 P(i,j)$	Measures the skewness of the GLCM (asymmetry)
6	Dissimilarity	$\sum_{i,j=1}^N  i-j P(i,j)$	Returns linear local variation boundary of the gray levels in a texture
7	Energy (uniformity)	$\sqrt{\sum_{i,j=1}^N P(i,j)^2}$	Returns the sum of squared elements in the GLCM.
8	Entropy	$-\sum_{i,j=1}^N P(i,j) \log P(i,j)$	Returns the degree of randomness and complexity of a texture
9	Homogeneity	$\sum_{i,j=1}^N P(i,j)/1+ i-j $	Returns the distribution consistency of the GLCM elements
10	Maximum probability	$\max_{i,j} P(i,j)$	Returns the largest $P(i,j)$ value found within the window
11	Sum of squares (variance)	$\sum_{i,j=1}^N (i-\mu)^2 P(i,j)$	Measures the dispersion of the values around the mean
12	Sum average	$\sum_{i=0}^{2G} iP_{x+y}$	Measures the mean value of Px and Py
13	Sum variance	$\sum_{i=1}^{2G} (i - (\text{sum entropy}))^2 P_{x+y}(i)$	Measures the variance of sum entropy
14	Sum entropy	$-\sum_{i=0}^{2G} P_{x+y}(i) \log(P_{x+y}(i))$	Measures the sum of Px and Py entropy
15	Difference variance	$\text{Variance} P_{(x-y)}$	Measures the variance difference between two values
16	Difference entropy	$-\sum_{i=0}^G P_{x-y}(i) \log(P_{x-y}(i))$	Measures the difference of Px and Py entropy
17	Information measure of correlation (1)	$\frac{HXY-HY1}{\max\{HX,HY\}}$	Measures the linear dependency of gray levels on those of neighboring pixels.
18	Information measure of correlation (2)	$(1 - \exp[-2(HXY2 - HXY)])^{0.5}$	Measures the linear dependency of gray levels on those of neighboring pixels.
19	Inverse difference normalized (INN)	$\sum_{i,j=1}^G \frac{1}{1+ i-j ^2/G^2} P(i,j)$	Returns the linear normalized distribution consistency of the GLCM elements
20	Inverse difference moment normalized	$\sum_{i,j=1}^G \frac{1}{1+(i-j)^2/G^2} P(i,j)$	Returns the normalized distribution consistency of the GLCM elements

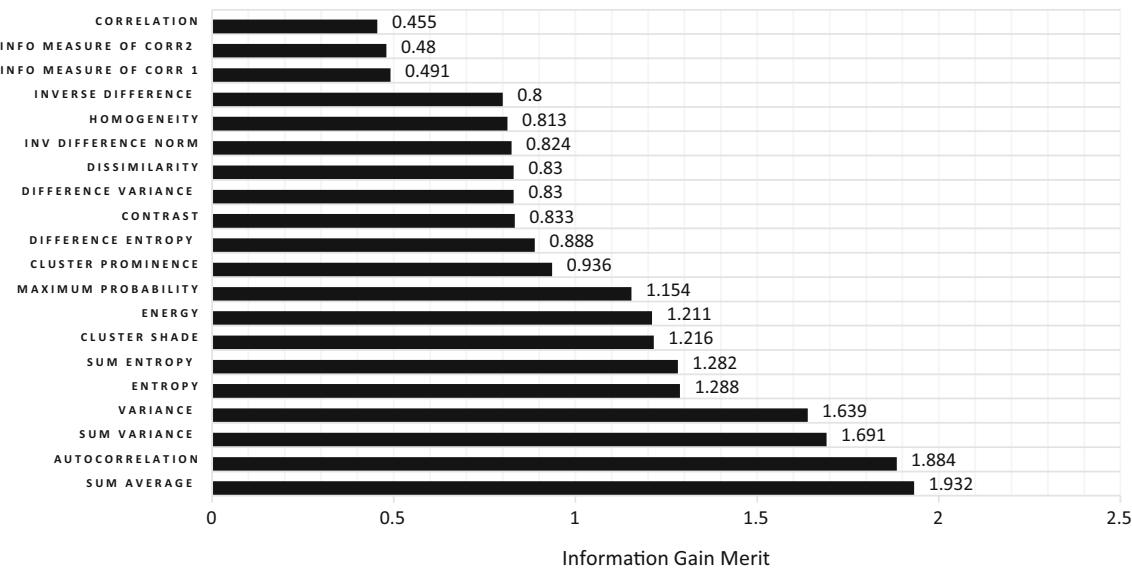
$p(i,j) = (i,j)$ th entry in normalized GLCM;  $P_x(i) = i$ th entry in the marginal probability matrix obtained by summing the rows of  $p(i,j)$ ;  $P_x(i) = \sum_{j=1}^G p(i,j)$ ;  $P_y(j) = j$ th entry in the marginal probability matrix obtained by summing the columns of  $p(i,j)$ ;  $P_y(j) = \sum_{i=1}^G p(i,j)$ ;  $G$  = number of gray levels in quantized image.  $\mu_x, \mu_y, \sigma_x, \sigma_y$  = means and standard deviation of  $p_x$  and  $p_y$ , respectively

technique populates the feature vector iteratively based on information gain merit and classification performance. At each iteration feature with the highest information gain merit will be added to the feature vector, respectively. The skin classification  $F$ -score of the feature vector measured in each iteration. Iteration stops if adding more features to the feature vector does not make any sensible improvement in  $F$ -score and that feature vector will be considered as optimal.

Figure 6 shows the  $F$ -score value of the optimal feature space at each iteration. Random Forest, SVM and MLP classifiers were used to evaluate each feature space subset. It can be seen that regardless of the classifier,  $F$ -score value reaches its global maxima when the first eight GLCM features with the highest information gain merit are involved in the exhaustive search. The  $F$ -score value of the optimal feature space subset increases dramatically

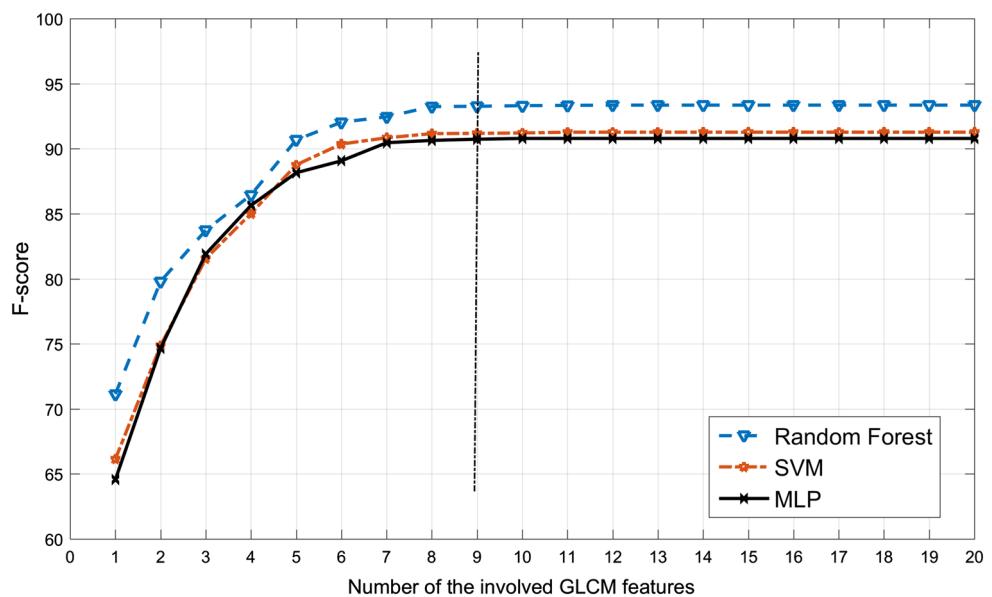
as the number of GLCM features increases from 1 to 9 features. The  $F$ -score remains relatively constant as the number of GLCM features is increased from 9 to 12 features. However, as we continued to involve more GLCM features into the feature vector subset, it was observed that the supervised exhaustive search mechanism did not incorporate the new GLCM features into the optimal solution. This might be due to the relatively low information gain merit of these features, which basically increase the noise and correlation.

Correspondingly, to avoid unnecessary computational overload, only the first nine GLCM features with the highest information gain merit including sum average, autocorrelation, sum variance, variance, entropy, sum entropy, cluster shade, energy and maximum probability were involved in finding the optimal feature space subset. By using the incremental exhaustive search, the search



**Fig. 5** Information gain merit for GLCM features

**Fig. 6** *F*-score of the optimal feature space subset as the number of the involved GLCM features increases, based on the information gain merit in Fig. 5



space for the optimal feature space subset has been significantly reduced without any compromise in accuracy of the results.

Table 2 shows the classification measures comparison between the original 140 dimension GLCM feature space versus the dimensionally reduced feature space, using the supervised incremental exhaustive search across the different classifiers, Random Forest, SVM and MLP, and different quantization levels, which were 8, 16, 32, 64, 128 and 256 gray levels. The *F*-score value has been used as the figure of merit in this comparison. According to Table 2, the highest *F*-score of 93.24 was observed in the dimensionally reduced feature space of 17 features in 64 gray-

level texture patches using the Random Forest classifier. The features in this feature space include 064\_05\_12, 064\_04\_01, 064\_05\_01, 064\_06\_12, 064\_06\_13, 064\_05\_11, 064\_06\_11, 064\_03\_08, 064\_01\_14, 064\_03\_05, 064\_07\_12, 064\_06\_01, 064\_07\_13, 064\_04\_10, 064\_02\_14, 064\_07\_11, 064\_04\_08, which can be interpreted using the instructions provided earlier. Table 2 shows that majority of the dimensionally reduced feature spaces have relatively similar or even higher *F*-scores compared to the original 140-dimension feature space, which indicates the presence of many noisy, redundant, correlated and irrelevant features in the original feature space.

From the quantization level standpoint presented in Fig. 7, it can be observed that *F*-score value dramatically rises as the quantization level increased from 8 to 64 gray levels. 64 gray levels consistently deliver the highest *F*-score value across different dimensionally reduced feature spaces, as well as the original 140-dimensional feature space. Higher numbers of gray levels (128 and 256) introduce arbitrary noise patterns to the skin texture and drop the classification performance, while lesser numbers of gray levels do not provide a comprehensive picture of the actual natural skin texture. Figure 7 also indicates that, by a considerable margin, the feature space which used Random Forest classifier consistently outperformed other feature spaces across the entire quantization level range, which implies the decent performance of Random Forest for skin texture classification. The SVM and MLP classifiers have relatively analogous performance in terms of skin texture classification across different gray levels.

Figure 8 compares the ROC curves of skin patch datasets between the original and dimensionally reduced

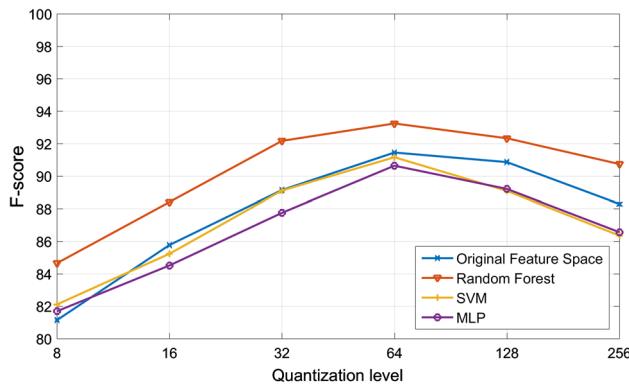
feature spaces. Due to the decent performance of 64 gray levels in skin texture analysis (Fig. 7), only this quantization level was used for sake of this experiment. The ROC curves show the tradeoff between FPR and TPR at various thresholds. Visual inspection of the ROC curves in Fig. 8 roughly indicates that dimensionally reduced feature space by Random Forest outperforms other feature spaces in terms of Area Under Curve (AUC). This feature space achieved AUC of 0.988, which is pleasing for the purpose of this research. Dimensionally reduced feature space by the MLP classifier had the least desirable performance in this experiment, although that does not imply the ineffectiveness of MLP in general terms. The original 140-dimensional feature space is the second best performer in this comparison; however, the computational complexity of this high-dimensional feature space makes it infeasible to be used in the real world.

Figure 9 shows the elapsed time (s) for testing a single  $150 \times 150$  texture patch taken from our in-house dataset using different dimensionally reduced and original feature spaces (shown in Table 2) across different quantization

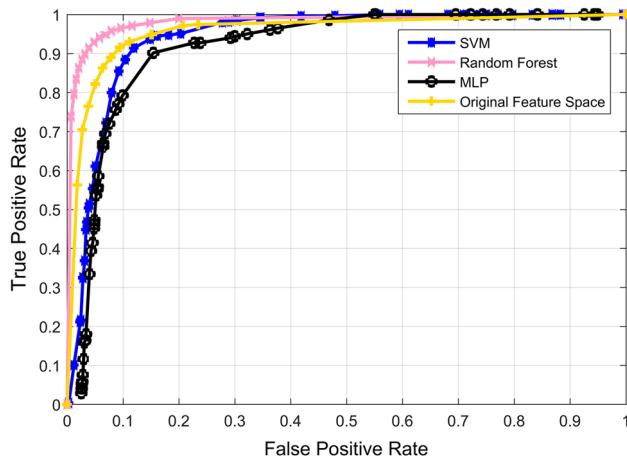
**Table 2** Performance analysis of the original feature space versus the dimensionally reduced feature space, using the incremental exhaustive feature selection and the supervised classifiers Random Forest, Support Vector Machine and Multilayer Perceptron

Methods	Quantization level	Dimensions	<i>F</i> -score	Accuracy	TPR	FPR
Original feature space—Random Forest	008 GL	140	81.14	82.16	81.96	15.70
	016 GL	140	85.77	85.91	85.84	11.02
	032 GL	140	89.14	89.68	88.42	6.89
	064 GL	140	91.45	91.86	90.07	5.83
	128 GL	140	90.86	91.07	90.71	6.12
	256 GL	140	88.29	88.77	87.24	8.36
Exhaustive feature selection—Random Forest	008 GL	12	84.65	84.82	84.81	12.23
	016 GL	14	88.41	88.93	87.65	8.39
	032 GL	14	92.17	92.68	91.41	3.49
	<b>064 GL</b>	<b>17</b>	<b>93.24</b>	<b>93.81</b>	<b>92.66</b>	<b>2.41</b>
	128 GL	20	92.33	92.97	91.91	3.32
	256 GL	21	90.74	91.16	90.55	6.19
Exhaustive feature selection—SVM	008 GL	16	82.11	82.01	81.63	15.42
	016 GL	18	85.23	84.65	85.54	15.19
	032 GL	18	89.11	88.92	89.30	11.15
	064 GL	20	91.16	91.44	89.67	5.08
	128 GL	23	89.09	89.30	89.11	7.94
	256 GL	27	86.35	86.91	86.14	10.03
Exhaustive feature selection—MLP	008 GL	15	81.69	82.06	81.41	16.05
	016 GL	15	84.51	84.27	84.08	13.71
	032 GL	18	87.75	87.54	86.88	11.02
	064 GL	19	90.64	90.65	90.67	9.37
	128 GL	21	89.21	89.43	88.96	10.98
	256 GL	22	86.56	86.74	86.39	11.55

Significant results are in bold

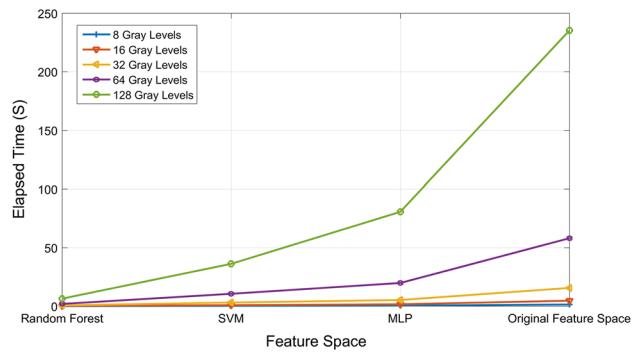


**Fig. 7** Comparison of  $F$ -score between the original feature space and dimensionally reduced feature space across different quantization values (gray levels)



**Fig. 8** Comparison of the ROC curves between the original feature space and dimensionally reduced feature spaces. A quantization level of 64 gray levels was used for sake of this experiment

levels including 8, 16, 32, 64 and 128. Figure 9 shows that Random Forest, by a relatively considerable margin, is the fastest classification technique in this study. The elapsed time of this classifier remains fairly reasonable across various quantization levels. SVM is the second fastest classifier in this comparison. Apart from the 128 gray levels, this classifier has relatively decent speeds across different quantization levels. Moving to the MLP classifier, it can be seen that elapsed time has been exponentially increased, especially at 64 and 128 gray levels. For instance, it takes roughly 20 s for this classifier to test a single 64 gray-level  $150 \times 150$  texture patch. Lastly, Fig. 9 implies that texture analysis using the original 140-dimension feature space is literally impractical and takes huge amount of time; moreover, as we have observed in Table 2, due to the presence of noisy and redundant features, this feature space is unable to generate any notable result.



**Fig. 9** Elapsed time (s) for testing a single  $150 \times 150$  texture patch using different dimensionally reduced and original feature spaces across different quantization levels

#### 4.2 Evaluation

In order to identify the effectiveness of the proposed skin texture model in the real-world scenario and measure its robustness, the proposed model was placed in a pixelwise skin detection as an auxiliary technique. The ECU dataset was used for this experiment [12]. The ECU dataset with over 4000 images can be an ideal representation of human skin which simulates variety of the real-world situations. Images in this dataset were acquired in uncontrolled lighting conditions and comprise various skin color hues and tones. This dataset also consists of image with skin-like color objects in the image background which further challenge skin segmentation. Therefore, this dataset can be an ideal choice to evaluate and compare the performance of the proposed texture model in the real world and test its robustness and repeatability. The performance results were compared to the basic pixelwise model to identify how much improvement could be achieved in pixelwise skin detection using the proposed skin texture model. Due to relatively decent performance of YCbCr in skin detection, this color space was employed in this experiment. The feature vector consisted of 20 features in total, comprising 17 textual features, 064\_05\_12, 064\_04\_01, 064\_05\_01, 064\_06\_12, 064\_06\_13, 064\_05\_11, 064\_06\_11, 064\_03\_08, 064\_01\_14, 064\_03\_05, 064\_07\_12, 064\_06\_01, 064\_07\_13, 064\_04\_10, 064\_02\_14, 064\_07\_11, 064\_04\_08, which had the highest performance in our texture evaluation (Table 2) and three color features, which were the Y, Cb and Cr color components. In the case of the basic pixelwise model, the feature vector consisted of only three color components. The Random Forest classifier was used to evaluate the performance of skin detection models in both cases. This study used the same parameters for Random Forest as used by Khan et al. [39].

As mentioned earlier, the ECU dataset has been used in this experiment for training and testing purposes. Training

and testing subdatasets were formed by splitting the dataset into two distinct complementary subsets in the ratio of 60% (2400 images) for training subsets and 40% (1600 images) for testing subsets. The ratio between training and testing subsets strongly depends on the size, variance and statistical distribution of the data. We identified a 60/40 ratio which has been practiced by many researchers [61, 62], which depicts nearly all possible patterns and variations in the entire input data in the training subset while avoiding arbitrary and bias results in the testing stage. Additionally, to further improve the reliability and trustworthiness of the experiments, a tenfold stratified cross-validation technique was employed to evaluate the performance of the proposed model.

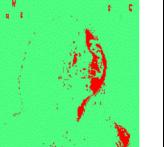
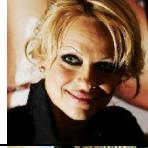
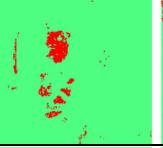
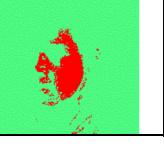
Table 3 compares the performance of the basic pixelwise skin detection model versus texture-assisted pixelwise skin detection. The *F*-score is used as the figure of merit in this experiment. According to Table 3, there is only a slight drop (~0.5%) in the majority of the measures in the testing set as compared to training set. However, in the basic pixelwise model there is a wider gap between the training and testing measures. This indicates the relatively high generalization ability and consistency of the proposed model in skin detection. The experimental results also indicate that there is only a negligible drop in some of the measures using cross-validation technique as compared to validation using a separate test set. A large amount of input data increases the repeatability and reliability of the experiments regardless of the validation technique employed. Texture-assisted pixelwise skin detection outperformed the basic pixelwise skin detection by relatively significant margin. The basic pixelwise skin detection generated *F*-score of 80.43 on the testing set; however, the texture-assisted pixelwise skin detection, with over 11% of improvement, generated *F*-score of 91.98, which is satisfactory considering the challenging scenario in the ECU dataset. In terms of FPR, the texture-assisted pixelwise skin detection significantly reduced the false detection rate from 16.36 to 5.84%. It can be concluded that the proposed skin texture model had a significant positive impact on pixelwise skin detection performance.

In terms of robustness, it can be implied that the proposed skin texture model is fairly robust and maintains its performance and effectiveness once it is being tested on a new independent data or a real-world scenario. According to Table 3, despite the fact that the training and testing subsets are completely different, the proposed texture model does not deteriorate in the testing stage and roughly generates similar results and errors in both training and testing subsets. Table 3 also indicates that the results from the cross-validation technique are roughly in par with the results using a separate test set. The consistent performance of the proposed model across the training, testing and cross-validation stages, on the one hand, and the comprehensiveness of the ECU dataset (mimics various real-world situations) on the other hand imply the robustness of the proposed model. Although the quantitative results express the proposed model's performance accurately, the qualitative results demonstrate a more tangible, real-world experience of the model's performance. Figure 11 qualitatively compares the performance of the basic pixelwise skin detection model versus texture-assisted pixelwise skin detection. The images in this figure which are selected from the ECU dataset containing skin-like texture patches such as sand or wood in the background demonstrate the proposed model's performance in challenging situations. Ground truth masked images are also provided as a reference, to give an impression of an ideal skin segmentation. As with the quantitative evaluation, the Random Forest classifier has been used to evaluate the performance of skin detection models in both cases. Comparing the results from basic and texture-assisted pixelwise skin detection with provided ground truth indicates that the proposed GLCM texture features significantly improve the pixelwise skin detection performance. The results from the texture-assisted model have noticeably fewer false detections as well as a higher true detection value. The basic pixelwise skin detection model which only relies on color information (in our case Y, Cb and Cr color components) seems to struggle with skin-like texture patches in the image background. However, the texture-assisted model delivers significantly superior performance in skin detection, due to use of texture features.

**Table 3** Performance comparison of the basic pixelwise skin detection model against texture-assisted pixelwise skin detection (the proposed model)

Methods	Training set			Testing set			Tenfold cross-validation		
	<i>F</i> -score	Accuracy	FPR	<i>F</i> -score	Accuracy	FPR	<i>F</i> -score	Accuracy	FPR
Basic pixelwise skin detection	82.67	84.15	13.81	80.43	81.77	16.36	80.31	81.54	16.49
Texture-assisted pixelwise skin detection (proposed model)	92.41	93.09	5.15	91.98	92.79	5.84	91.90	92.63	5.93

The ECU dataset and Random Forest classifier were used to evaluate both models

Condition	Original Image			Segmented Image		
Overexposed / Underexposed skin regions						
Extremely poor or colored light source						
Low quality and small image						

**Fig. 10** Some examples where the proposed model failed to properly segment the skin regions. In segmented images, *red color* represents the regions which predicted as skin while *green color* shows the regions predicted as non-skin (color figure online)

Despite the overall satisfactory performance of the proposed skin texture model, we have realized that in some minor circumstances the proposed model failed to properly segment the skin regions. Our observations on the ECU dataset segmentation results showed that the proposed model had difficulty in segmenting the skin region in highly overexposed or underexposed images, small and low-quality images and extremely poor or colored lighting conditions. In the majority of these circumstances, skin texture does not hold its statistical attributes, as well as its color spectrum, which degrades the segmentation performance. Figure 10 shows some extreme examples for which the proposed model failed to properly segment the skin regions.

## 5 Comparison

This section presents a performance comparison between the proposed texture-assisted pixelwise skin detection and some notable state-of-the-art skin texture detection techniques. Comparison between approaches is valid and unbiased only if all approaches were trained and tested with identical input data, experiment environments and factors. Hence, the ECU benchmark dataset, which is explained in Sect. 3, was used in this comparison for all participant studies. In this comparison, training and testing subdatasets were formed by splitting the ECU Dataset into two distinct complementary subsets with ratio of 60% (2400 images) for training subsets and 40% (1600 images) for the testing subsets. All the approaches participating in this comparison were carefully implemented according to the techniques and parameters mentioned in their work.

Table 4 provides a comparison measures between the proposed texture-assisted skin detection model and some well-known state-of-the-art skin detection models. Brief details of feature extraction techniques as well as the employed detection method are also provided in this table. Performance measures in this comparison were accuracy, *F*-score and FPR. The comparison results show that the proposed texture-assisted skin detection method, with accuracy, *F*-score and FPR of 92.79, 91.98 and 5.84, respectively, outperformed the other skin detection models in this comparison. Although the ECU dataset contains images with uncontrolled lighting condition and skin-like backgrounds, the proposed model managed to generate satisfactory FPR of 5.84%. The approaches of Cheddad et al. [63] and Zaidan et al. [36], with FPR of 8.14 and 9.77, also generated promising results. In terms of *F*-score, which is an all-round performance measure, once again the proposed model managed to achieve *F*-score of 91.98, which is superior to any other model tested in this comparison. Cheddad et al. [63] with *F*-score of 90.97 marginally underperformed the proposed model. One major drawback in the approach of Cheddad et al. [63] is the assumption of face presence in the image. Their proposed technique relies on face detection in order to make a better estimation of skin color; however, once the given image does not represent any human face, this technique fails completely. The color-based model proposed by Zaidan et al. [36] generates relatively good results, considering that they have not employed any textural analysis techniques. The qualitative comparison results between the proposed texture-assisted model and the methods of Zaidan et al. [36], Kawulok et al. [64] and Cheddad et al. [63] in Fig. 11 roughly show the advantage

**Table 4** Comparison of the proposed texture-assisted pixelwise skin detection with state-of-the-art skin detection techniques

References	Feature extraction technique	Detection method	Accuracy	F-score	FPR
[36]	Skin segmentation using the grouping histogram and segment adjacent-nested techniques	Backpropagation Neural network classifier	<b>89.13</b>	<b>88.56</b>	<b>9.77</b>
[64]	Skin segmentation using textural features from the skin probability	Linear discriminant analysis used to classify the skin regions	<b>86.45</b>	<b>86.12</b>	<b>16.45</b>
[65]	A region growing skin segmentation techniques using the initial seeds from color distance map	Static Threshold based classification	75.18	72.70	32.96
[66]	Skin segmentation based on mean and standard deviation color moments, color features as well as GLCM features including homogeneity, contrast and correlation	Neural network classification technique	81.02	80.66	23.89
[67]	Texture descriptors including uniformity, standard deviation, skewness, kurtosis, smoothness and entropy	Multilayer Perceptron neural network combined with the k-means clustering techniques	<b>86.16</b>	<b>84.65</b>	<b>17.33</b>
[68]	Cb, Cr chrominance components and GLCM skin texture features including entropy, energy, contrast, homogeneity	feed forward backpropagation neural network	71.02	69.64	35.91
[63]	Threshold-based skin probability map using one-dimensional illumination-based color space	Expectation maximization/ Gaussian mixture model	<b>90.13</b>	<b>90.97</b>	<b>8.14</b>
[69]	Color histogram based on adoptive Cb, Cr color component-based initial face or hand detection	Threshold-based skin segmentation	73.55	71.76	27.63
[70]	Two-stage skin detection based on histogram information in YCbCr color space and texture information using Daubechies Wavelet transformation	Gaussian mixture model	79.04	80.55	22.18
[71]	Gray statistics features for texture features and implements skin color detection combining with HSV color space.	Static threshold-based classification	82.96	81.76	19.69
[72]	Color features in RGB color space along with texture features such as standard deviation, range and entropy	Backpropagation artificial neural networks	75.34	75.95	22.67
[73]	Skin detection using a modified probability histogram techniques and Gaussian mixture model technique along with texture features using non-subsampled contourlet coefficients	Multilayer Perceptron classification technique	83.67	83.42	17.64
[74]	Skin detection using combination of color features (CbCr), texture features (Gabor filter) as well as watershed segmentation	Skin probability map	77.18	78.24	21.28
Proposed texture-assisted pixelwise skin detection	YCbCr color components alongside 17 optimized GLCM skin texture features	Random Forest decision tree	<b>92.79</b>	<b>91.98</b>	<b>5.84</b>

Significant results are in bold

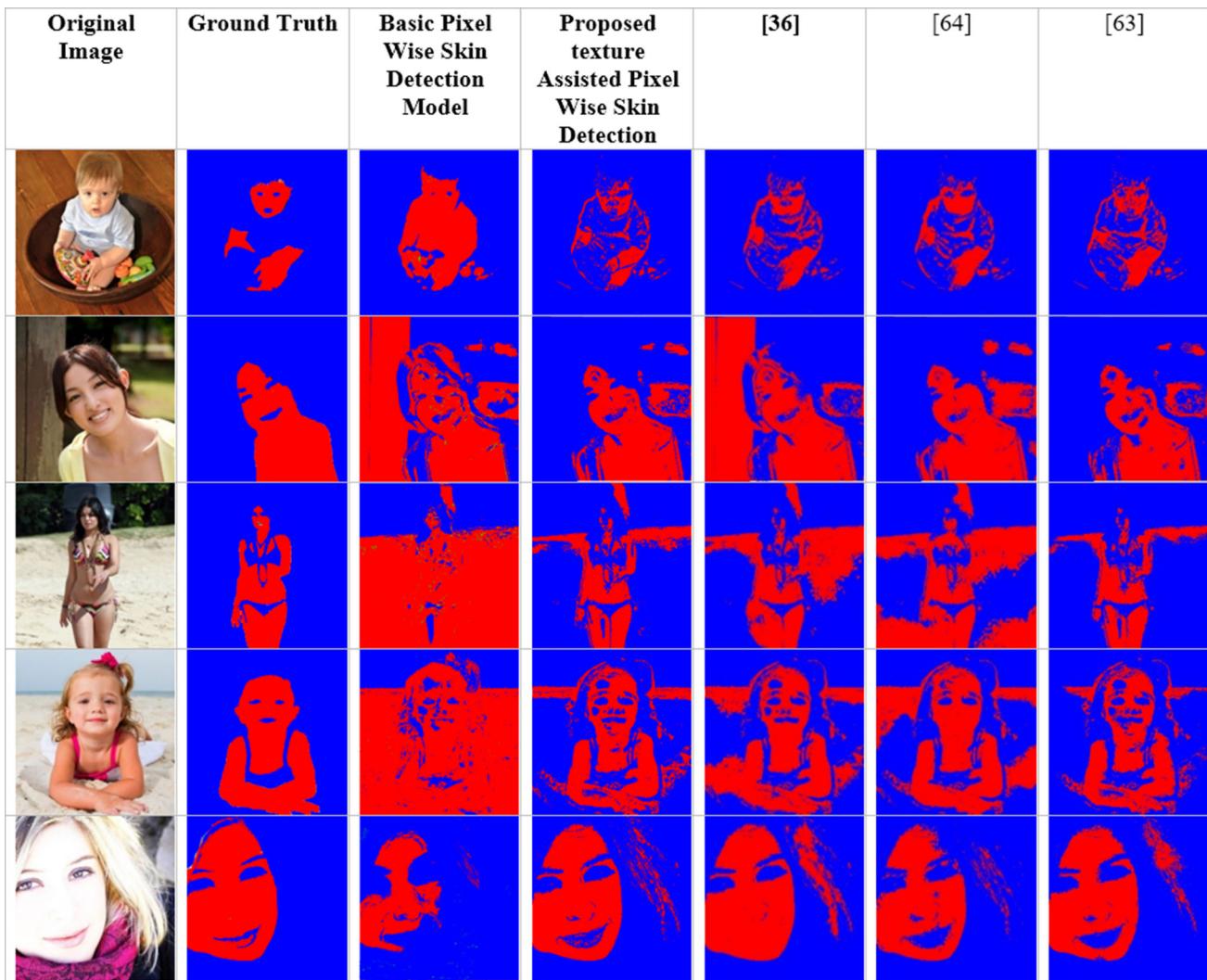
The ECU benchmark dataset was used for the purpose of this comparison

of our proposed model and testify and support the comparison results in Table 4.

## 6 Conclusion

Methods relying on the color information solely have proved to be incapable of accurately differentiating the skin-like color regions from the actual skin regions in pixelwise skin detection. Texture features have been devised to counter this issue by considering the statistical and spatial relations among the group of neighboring pixels

in the image instead of relying only on the color information of individual pixels. Among the various techniques that have been utilized for texture analysis, GLCM has proved to be well suited for texture analysis of a natural object such as human skin. Numerous studies have attempted to analyze human skin texture through GLCM features; however, the majority of these used arbitrary values to set up GLCM parameters. Thus, improper or arbitrary adoption of these parameters led to the ineffectiveness of the existing GLCM techniques in skin detection. Accordingly, this research has attempted to address this issue by optimizing the GLCM parameters including



**Fig. 11** Qualitative comparison of the basic pixelwise skin detection model versus texture-assisted pixelwise skin detection and two other state-of-the-art skin detection techniques. The red color represents the

predicted skin patches while the blue color denotes the non-skin areas of the image (color figure online)

quantization level, displacement magnitudes, displacement direction and GLCM features using an exhaustive supervised search and machine learning techniques to improve the effectiveness of GLCM texture features for skin detection. It has been discovered that GLCM features including sum average, autocorrelation, sum variance, variance, entropy, sum entropy, cluster shade, energy and maximum probability have high enough information gain merit and carry enough discriminative information to be used for skin texture classification. Furthermore, it has been identified that displacement direction has minimal impact in skin texture analysis using GLCM. We also found that textural information in photographic images gradually fades away as the displacement magnitude is increased from 1 to 7 pixels. In addition, this research has been identified that the most proper quantization in skin texture analysis using GLCM is 64. The exhaustive

incremental supervised search discovered that a combination of 17 GLCM features including 064\_05\_12, 064\_04\_01, 064\_05\_01, 064\_06\_12, 064\_06\_13, 064\_05\_11, 064\_06\_11, 064\_03\_08, 064\_01\_14, 064\_03\_05, 064\_07\_12, 064\_06\_01, 064\_07\_13, 064\_04\_10, 064\_02\_14, 064\_07\_11, 064\_04\_08 forms the most effective GLCM texture feature space for skin detection. In terms of skin classification techniques, this study identified that Random Forest classifier has superior performance in skin detection compared to MLP and SVM classifiers. Our experiments showed that pixelwise skin detection accuracy can be significantly improved by embedding the proposed optimized GLCM skin texture feature as an auxiliary feature set into the feature space.

The findings of this research are expected to have a positive impact on various applications of skin detection such as face recognition, human detection, hand gesture

recognition and skin disorder and lesion recognition. These findings also provide a better understanding of skin texture behavior in photographic images. Apart from skin detection and its applications, the proposed methodology in finding the optimal GLCM features and parameters can be applied to different image segmentation disciplines, such as computed tomography image and radar images. Despite the significant progress in skin texture analysis that has been achieved so far, this research uncovered new possible areas for further research. The proposed skin texture model has been evaluated using the images in the ECU dataset. Other skin benchmarking datasets can be used to evaluate the performance of the proposed skin texture model, to increase its reliability and trustworthiness. Moreover, the impact of illumination variation and skin color hue on skin texture can be investigated to improve the robustness of skin detection against these issues. This research showed that integration of an optimal set of textural and color features can bring significant improvement in pixelwise skin detection; nevertheless, we believe that there is still room for optimizing the GLCM skin textural features.

**Acknowledgements** The authors would like to acknowledge ministry of higher education (MOHE) and Universiti Teknologi Malaysia (UTM) for supporting this research under Research University Grant (RUG) vote 10J28 and science fund Grant (MOSTI) vote 01-01-06-SF1167.

#### Compliance with ethical standards

**Conflict of interest** We would like to state that there is no conflict of interest to declare and corresponding author shall bear full responsibility for the submission.

#### References

- Zhang S, Yang H, Singh L (2014) Increased information leakage from text. In: Proceedings of CEUR workshop, vol 1225, no 1, pp 41–42
- Brown DA, Craw I, Lewthwaite J (2001) A SOM based approach to skin detection with application in real time systems. BMVC 1:491–500
- Tuceryan M, Jain AK (1993) Texture analysis. In: The handbook of pattern recognition and computer vision, pp 207–248
- Lloyd K, Rosin PL, Marshall D, Moore SC (2016) Detecting violent crowds using temporal analysis of GLCM texture. arXiv preprint [arXiv:1605.05106](https://arxiv.org/abs/1605.05106)
- Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3(6):610–621
- Gonzalez RC, Woods RE (2007) Digital image processing, 3rd edn. Prentice Hall, Upper Saddle River
- Vezhnevets V, Sazonov V, Andreeva A (2003) A survey on pixel-based skin color detection techniques. In: Proceedings of graphicon, vol 85, pp 85–92
- Gadelmawla ES (2004) A vision system for surface roughness characterization using the gray level co-occurrence matrix. NDT E Int 37(7):577–588
- Kuffer M, Pfeffer K, Sliuzas R, Baud I (2016) Extraction of slum areas from VHR imagery using GLCM variance. IEEE J Select Top Appl Earth Obs Remote Sens 9(5):1830–1840
- Daliman S, Rahman SA, Busu I (2014) Segmentation of oil palm area based on GLCM- SVM and NDVI. In: Region 10 symposium IEEE, pp 645–650
- Renzetti FR, Zortea L (2011) Use of a gray level co-occurrence matrix to characterize duplex stainless steel phases microstructure. Frattura ed Integrità Strutturale 16:43–51
- Phung SL, Bouzerdoum A, Chai D (2005) Skin segmentation using color pixel classification: analysis and comparison. IEEE Trans Pattern Anal Mach Intell 27(1):148–154
- Soh L, Tsatsoulis C (1999) Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Geosci Remote Sens 37(2):780–795
- Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens 28(1):45–62
- Zhang D, Zhao M, Zhou Z, Pan S (2013) Characterization of wire rope defects with gray level co-occurrence matrix of magnetic flux leakage images. J Nondestruct Eval 32(1):37–43
- Ou X, Pan W, Xiao P (2014) In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). Int J Pharm 460(1–2):28–32
- Xian GM (2010) An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. Expert Syst Appl 37(10):6737–6741
- Maurya R, Singh SK, Maurya AK, Kumar A (2014) GLCM and multi class support vector machine based automated skin cancer classification. In: IEEE international conference on computing for sustainable global development (INDIACom), pp 444–447
- Zhu H, Zhou S, Wang J, Yin Z (2007) An algorithm of pornographic image detection. In: Fourth international conference image and graphics, 2007, ICIG 2007, pp 801–804
- Liu Y, Zhang H, Li P (2011) Research on SVM-based MRI image segmentation. J China Univ Posts Telecommun 18(December):129–132
- Wang X, Zhang X, Yao J (2011) Skin color detection under complex background. In: International conference on mechatronic science, electric engineering and computer, pp 1985–1988
- Jeniva S (2015) An efficient skin lesion segmentation analysis using statistical texture distinctiveness. Int J Adv Res Trends Eng Technol 3777:111–116
- Pang H, Chen T, Wang X, Chang Z, Shao S, Zhao J (2017) Quantitative evaluation methods of skin condition based on texture feature parameters. Saudi J Biol Sci 24(3):514–518
- Gómez W, Pereira WCA, Infantosi AFC (2012) Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. IEEE Trans Med Imaging 31(10):1889–1899
- De La Casa Almeida M, Serrano CS, Rejano JJ, Díaz JR, Lugo MLB, Roldán JRR (2015) Reliability of texture analysis using co-occurrence matrices (glcm) on photographic image in the assessment of cellulite in a Spanish population. J Eur Acad Dermatol Venereol 29(2):315–324
- Zhang X, Weng C, Yu B, Li H (2014) In-vivo differentiation of photo-aged epidermis skin by texture-based classification. In: SPIE/COS photonics Asia international society for optics and photonics
- Aswin RB, Jaleel JA, Salim S (2014) Hybrid genetic algorithm—artificial neural network classifier for skin cancer detection. In: IEEE international conference on control, instrumentation, communication and computational technologies (ICCI CCT), pp 1304–1309

28. Das N, Pal A, Mazumder S, Sarkar S, Gangopadhyay D, Nasipuri M (2013) An SVM based skin disease identification using local binary patterns. In: IEEE third international conference on Advances in computing and communications (ICACC), pp 208–211
29. Pengyu N, Jie H (2013) Pornographic image filtering method based on human key parts. In: Proceedings of the international conference on information technology and software engineering, vol 212, pp 677–688
30. Wang YWY, Wu XWX, Yang LY (2010) Sensitive body image detection technology based on skin color and texture cues. In: 3rd International congress on image signal processing (CISP), vol 6, pp 2661–2664
31. Yue J, Li Z, Liu L, Fu Z (2011) Content-based image retrieval using color and texture fused features. *Math Comput Model* 54(3–4):1121–1127
32. El-Khamy SE, Abdel Alim OA, Saïi MM (2001) Neural network face recognition using statistical feature and skin texture parameters. In: Proceedings of the eighteenth national on radio science conference NRSC2001, pp 233–240
33. Al Abbadi NK, Dahir NS, Abd Alkareem Z (2013) Skin texture recognition using neural networks. In: 2008 International Arab conference on information technology (ACIT 2008), pp 3–6
34. Schwartz WR, Kembhavi A, Harwood D, Davis LS (2009) Human detection using partial least squares analysis. In: IEEE 12th international conference on computer vision, pp 24–31
35. Clausi DA, Jernigan ME (1998) A fast method to determine co-occurrence texture features. *IEEE Trans Geosci Remote Sens* 36(1):298–300
36. Zaidan AA, Ahmad NN, Abdul Karim H, Larbani M, Zaidan BB, Sali A (2014) Image skin segmentation based on multi-agent learning Bayesian and neural network. *Eng Appl Artif Intell* 32:136–150
37. Cao X, Liu H (2012) A skin detection algorithm based on Bayes decision in the YCbCr color space. *Appl Mech Mater* 126:672–676
38. Khan R, Hanbury A, Stöttinger J, Bais A (2012) Color based skin classification. *Pattern Recognit Lett* 33(2):157–163
39. Khan R, Hanbury A, Stoettinger J (2010) Skin detection: a random forest approach. In: IEEE 17th international conference on image processing, pp 4613–4616
40. Zuo H, Hu W, Wu O (2010) Patch-based skin color detection and its application to pornography image filtering. In: Proceedings of 19th international conference on world wide web, pp 1227–1228
41. Maktabdar Oghaz M, Maarof MA, Zainal A, Rohani MF, Yaghoubyan SH (2015) A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique. *PLoS ONE* 10(8):e0134828
42. Polpinij J, Chotthanom A, Sibunruang C, Chamchong R, Puangpronnitrag S (2006) Content-based text classifiers for pornographic web filtering. In: IEEE international conference on systems, man and cybernetics, pp 1481–1485
43. Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. *IEEE Trans Pattern Anal Mach Intell* 24(1):34–58
44. Phung SL, Bouzerdoum A, Chai D (2005) Skin segmentation using color pixel classification: analysis and comparison. *IEEE Trans Pattern Anal Mach Intell* 27(1):148–154
45. Lin C (2006) Face detection in non-uniform illumination conditions by using color and triangle-based approach. In: Proceedings of 9th conference on JCIS science, pp 4–7
46. Lee J-S, Kuo Y-M, Chung P-C, Chen E-L (2007) Naked image detection based on adaptive and extensible skin color model. *Pattern Recognit* 40(8):2261–2270
47. Romero-Lopez A, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In: Biomed engineering (NY), pp 49–54
48. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115–118
49. Liao H (2016) A deep learning approach to universal skin disease classification. Department of Computer Science, University of Rochester, pp 1–8
50. Jafari MH, Karimi N, Nasr-Esfahani E, Samavi S, Soroushmeir SMR, Ward K, Najarian K (2016) Skin lesion segmentation in clinical images using deep learning. In: 23rd international conference on pattern recognition, pp 337–342
51. Li Y, Esteva A, Kuprel B, Novoa R, Ko J, Thrun S (2016) Skin cancer detection and tracking using data synthesis and deep learning, pp 1–4. arXiv preprint [arXiv:1612.01074](https://arxiv.org/abs/1612.01074)
52. Liu Z, Luo P, Wang X, Tang X (2014) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 3730–3738
53. Bharati A, Singh R, Vatsa M, Bowyer KW (2016) Detecting facial retouching using supervised deep learning. *IEEE Trans Inf Forensics Secur* 11(9):1903–1913
54. Xing J, Li K, Hu W, Yuan C, Ling H (2016) Diagnosing deep learning models for high accuracy age estimation from a single image. *Pattern Recognit* 66:106–116
55. Perez M, Avila S, Moreira D, Moraes D, Testoni V, Valle E, Goldenstein S, Rocha A (2017) Video pornography detection through deep learning techniques and motion information. *Neurocomputing* 230:279–293
56. Zhang H, Cao X, Ho JKL, Chow TWS (2016) Object-level video advertising: an optimization framework. *IEEE Trans Ind Inf* 13(99):520–531
57. Zhang H, Li J, Ji Y, Yue H (2017) Understanding subtitles by character-level sequence-to-sequence learning. *IEEE Trans Ind Inform* 13(2):616–624
58. Kent JT (1983) Information gain and a general measure of correlation. *Biometrika* 70(1):163–173
59. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data mining: practical machine learning tools and techniques. Morgan Kaufmann, Burlington
60. Lee C, Lee GG (2006) Information gain and divergence-based feature selection for machine learning-based text categorization. *Inf Process Manage* 42(1):155–165
61. Shehab T, Farooq M (2013) Neural network cost estimating model for utility rehabilitation projects. *Eng Constr Archit Manag* 20(2):118–126
62. Pivezhandi M, Maybodi BM (2015) Statistical based neural network in human activity recognition. *Int J Comput Appl* 124(12):1–5
63. Cheddad A, Condell J, Curran K, Mc Kevitt P (2009) A skin tone detection algorithm for an adaptive approach to steganography. *Sig Process* 89:2465–2478
64. Kawulok M, Kawulok J, Nalepa J (2014) Spatial-based skin detection using discriminative skin-presence features. *Pattern Recognit Lett* 41:3–13
65. Abdullah-Al-Wadud M, Shoyaib M, Chae O (2008) A skin detection approach based on color distance map. *EURASIP J Adv Signal Process* 2008(1):814283
66. Dumitrescu CM, Dumitrasche I (2013) Human skin detection using texture information and vector processing techniques by neural networks. *Adv Intell Control Syst Comput Sci* 59–75
67. Al-Mohair HK, MohamadSaleh J, Suandi SA (2015) Hybrid human skin detection using neural network and k-means clustering technique. *Appl Soft Comput* 33:337–347
68. El Abbadi NK, Dahir N, Alkareem ZA (2013) Skin texture recognition using neural networks. arXiv Preprint [arXiv:1311.6049](https://arxiv.org/abs/1311.6049)
69. Bilal S, Akmeliawati R, Salami MJE, Shafie AA (2012) Dynamic approach for real-time skin detection. *J Real-Time Image Process* 10(2):1–15

70. Ng P, Pun C-M (2011) Skin color segmentation by texture feature extraction and k-mean clustering. In: Computational intelligence, communication systems and networks, pp 213–218
71. Lei Y, Xiaoyu W, Hui L, Dewei Z, Jun Z (2011) An algorithm of skin detection based on texture. In: 4th international congress on image and signal processing (CISP), pp 1822–1825
72. Taqa AY, Jalab HA (2010) Increasing the reliability of skin detectors. *Sci Res Essays* 5(17):2480–2490
73. Fotouhi M, Rohban MH, Kasaei S (2009) Skin detection using contourlet-based texture analysis. In: Fourth international conference on digital telecommunications ICDT'09, pp 367–372
74. Jiang Z, Yao M, Jiang W (2007) Skin detection using color, texture and space information. In: Fourth international conference on fuzzy systems and knowledge discovery FSKD, pp 366–370