An Automatic Identification System of Human Skin Irritation

Abdul Fadlil
Dept. of Electrical Engineering, Faculty of Industrial Technology, Ahmad Dahlan University
Jl. Prof. Dr. Soepomo, Janturan, Yogyakarta 55164, Tel. +62-274-379418
e-mail: fadlil3@yahoo.com

Abstrak

Karakterisasi kuantitatif untuk kulit manusia yang mengalami iritasi adalah penting tetapi merupakan pekerjaan yang susah dilakukan. Sekarang ini untuk mengidentifikasi kulit manusia masih dilakukan secara manual. Tentu saja, identifikasi secara manual dari iritasi kulit manusia ini dapat menjadi subyektif. Analisis iritasi kulit telah dapat dilakukan menggunakan pengujian biokimia, namun tidak sederhana. Pada penelitian ini telah dikembangkan suatu pendekatan baru sistem identifikasi iritasi kulit manusia secara otomatis berbasis pengenalan pola citra untuk mendapatkan keputusan kulit yang diuji merupakan kulit yang mengalami iritasi atau tidak. Rancangan sistem yang dikembangkan ini menggunakan metode ekstaksi ciri: gray level histogram (GLH) dan ciri tekstur gray level co-occurrence matrices (GLCM). Selanjutnya untuk proses klasifikasi menggunakan metode metrik jarak: Manhattan dan Euclidean dan jaringan syaraf kuantisasi vektor pembelajaran (LVQ-NN). Untuk mengevaluasi unjuk kerja sistem dilakukan dengan mengkombinasikan metode-metode pengekstaksi ciri dan pengklasifikasi. Hasil-hasil percobaan menunjukkan bahwa akurasi terbaik adalah 83,33% yang diperoleh tatkala rancangan sistem diimplementasikan menggunakan ciri GLH atau GLCM melalui pengklasifikasi LVQ-NN.

Kata kunci: *sistem identifikasi, pengenalan citra, iritasi kulit*

Abstract

Quantitative characterization of human skin irritation is important but it is difficult task to be done. Recently, an identification of human skin is still doing manually. Indeed, the identification of the human skin irritation sample can be very subjective. The analysis of the skin irritation could be conducted using biochemical test, but it is not simple. In this research, a new approach of an automatic human skin identification system based on image pattern recognition is developed to obtain a decision of sample test (whether it has irritation or not). This system design was developed using the following features extraction: gray level histogram (GLH) feature and texture gray level co-occurrence matrices (GLCM). Meanwhile, for a classification process, using the following distance metric: Manhattan distance and Euclidean distance, or learning vector quantization neural network (LVQ-NN). The combination between feature extractor and classifier methods proposed was used to evaluate the performance system. The experimental results show that the best accuracy for 83.33% was obtained when design system was implementated using GLH or GLCM features through LVQ-NN classifier.

Keywords: *identification system, image recognition, skin irritation*

1. Introduction

With the advancement in the computer and information technologies, computers have been used as a common and convenient equipment for computations and automation applications. Recently, identification of human skin is still doing manually. Furthermore, manual identification of the human skin irritation sample can be very subjective. The skin irritation analysis was conducted using a biochemical test [1]. In this research, it have been developed an automatic identification system for human skin irritation based on an image recognition.

Image recognition is one of the field of pattern recognition. The recognition problem is usually approached to one of two modes namely identification and verification tasks. For the skin irritation identification system, a probe skin image sample is presented and the system is identified by computing the similarities between the probe skin image sample and all gallery skin image samples in the database.

The developing of image recognition system is quite difficult because image is quite complex, and correspond to environment changes. Many image recognition techniques have been proposed and have shown significant promise, but robust image recognition is still difficult. In the literature, the fish recognition based on robust features extraction from size and shape measurements using neural network has been proposed [2]. Khalid, *et.al* have developed a robust design of recognition system for an intelligent wood species [3].

Image skin analysis can be done based on skin color [4] and texture [2], [3], [5], [6]. In this research, it have been analyzed the comparative performance of features and classifier methods. The two features are gray level histogram (GLH) features and gray level co-occurrence matrices (GLCM) and the three classifiers are Manhattan distance and Euclidean distance, and learning vector quantization neural network (LVQ-NN). Combination between feature and classifier methods have been proposed to evaluate the performance system. This paper has been organized as follows. The next section describes research methods which include the data collection and acquisition, preprocessing, feature extractor, and classifier. The results and analysis will be discussed in Section 3. The last section concludes the paper.

2. Research Method

The research methodology used for this measurement can be described through the following points: data collection and acquisition, feature extraction and classification. A block diagram of the identification system for human skin irritation is shown as in Figure 1.

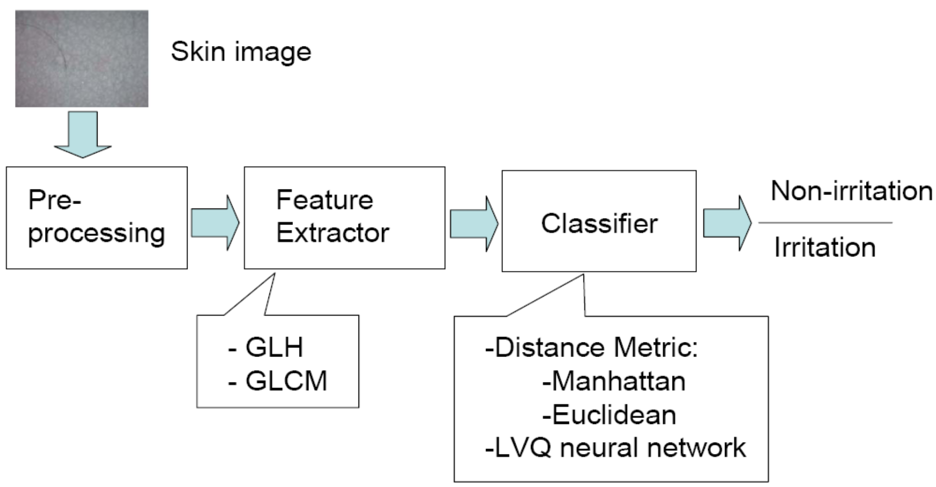

Figure 1. A block diagram of the identification system for human skin irritation**.**

Table 1. Training and test of the skin data sets

Skin image	Training	Test	Size	Format
Non-irritation	10	12	800x600	*.bmp
Irritation	10	12		

An example of the skin color images for non-irritation (Figure 2 a and b, respectively) and irritation (Figure 2 c and d, respectively).

2.1. Data collection and aquisition

The experimental system has been set up to acquire the images of the skin irritation or not using an industrial Handheld Digital Microscope Seri Dino-Lite Model AM-211. Data have been collected from five person and then the database was divided into two parts namely for a training phase and a testing phase. All data sets of the skin images for training and testing phase are shown in Table 1.

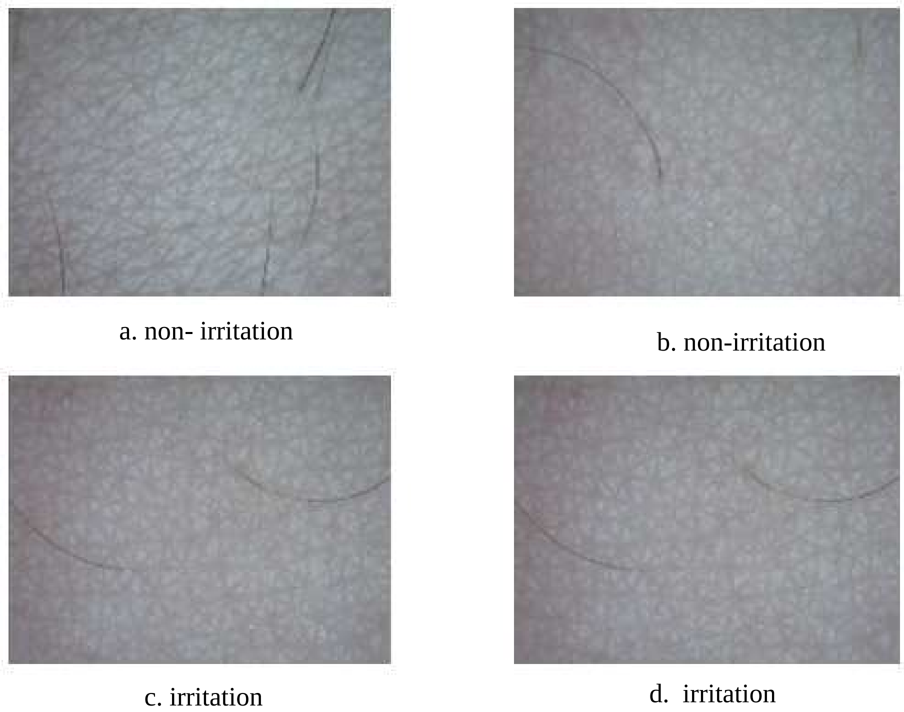

Figure 2. An example of the image skin true color or RGB

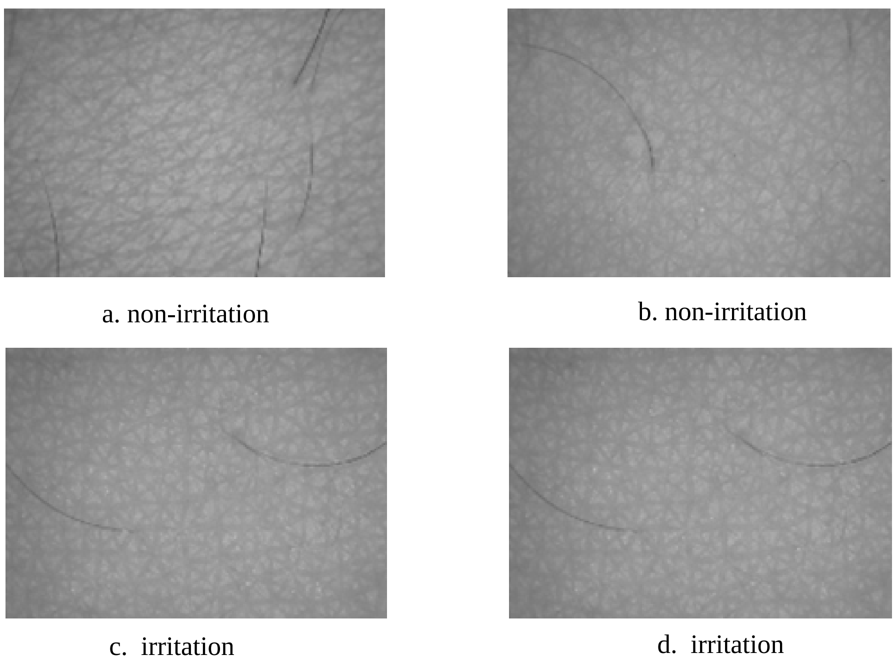

Figure 3. An example of the image skin grayscale

2.2. Pre-processing

In this work, the goal of the preprocessing is to covert a true color image to grayscale. The value of gray scale image is obtained from RBG by referring to the following formula:

$$\mathrm{W}_{\mathrm{grayscale}} = 0.5R + 0.3G + 0.2B \quad (1)$$

In the pre-processing phase, the skin image of true color will be converted to gray-scale image. The original images as shown in Figure 2 has been converted to gray level as shown in Figure 3.

2.3. GLH and GLCM Feature Extactor

In this research, it has been implemented two features models namely: a GLH and a GLCM. A gray-scale histogram shows how many pixels contained in each gray level (intensity level). Each pixel is a shade of gray, normally from 0 to 255. Therefore, skin image based on gray level histogram is a feature vector 256-dimension. The GLH normalized obtained by devide maximum value of histogram as features or input vector for classifier.

Texture analysis methods have been utilized in a variety of image recognition application. The famous texture analysis method is a GLCM, which was introduced by Haralick [7]. The GLCM features have been implemented in some applications such as fish recognition [2], wood species recognition [3], and defect detection [8]. The GLCM is a tabulation shows of how often the different combinations of pixel brightness value (gray levels). The GLCM describes here is the joint probability of occurrence of gray levels *a* and *b* for two pixels with a defined spatial relationship in an image. The spatial relationship is defined in terms of distance d and angle θ. If the texture is coarse and the distance d is small compared to the size of the texture elements, the pairs of points at distance d should have similar gray levels. Conversely, for a fine texture, if the distance d is comparable to the texture size, then the gray levels of points separated by distance d should often be quite different, so that the values in the GLCM should be spread out relatively uniform. From these matrices, a variety of features may be extracted. The matrices are constructed at a distance of d = 1 from four directions of angles $\theta=0^o$, 45^o, 90^o, 135^o, respectively. The matrices can be represented formally as follows [7]:

$$P_{0^o,d}(a,b) = |[(k,l),(m,n)] \in D : k-m=0, |l-n| = d| \quad (2)$$

$$P_{45^o,d}(a,b) = |[(k,l),(m,n)] \in D : (k-m, l-n) = (d,-d) or (-d,d)| \quad (3)$$

$$P_{90^o,d}(a,b) = |[(k,l),(m,n)] \in D : |k-m| = d, l-n=0| \quad (4)$$

$$P_{135^o,d}(a,b) = |[(k,l),(m,n)] \in D : (k-m, l-n) = (d,d) or (-d,-d)| \quad (5)$$

with *f(k; l) = a* and *f(m; n) = b* for all above equations.

The total features extracted from the GLCM approach from each skin sample are given as follows [7], [9]:

1. Energy or angular second moment:

$$\sum_{a,b} P_{\phi,d}^2(a,b) \quad (6)$$

2. Inertia or variance:

$$\sum_a \sum_b (a-b)^2 P_{\phi,d}(a,b) \quad (7)$$

3. Entropy:

$$\sum_{a,b} P_{\phi,d}(a,b) \log_2 P_{\phi,d}(a,b) \quad (8)$$

4. Homogeneity:

$$\sum_a \sum_b \frac{1}{1+(a+b)^2} P_{\phi,d}(a,b) \tag{9}$$

5. Maximum Probability:

$$\max_{a,b} P_{\phi,d}(a,b) \tag{10}$$

6. Contrast:

$$\sum |a-b|^k P_{\phi,d}^{\lambda}(a,b) \tag{11}$$

7. Inverse Difference moment:

$$\sum_{a,b;a\neq b} \frac{P_{\phi,d}^{\lambda}(a,b)}{|a-b|^k} \tag{12}$$

8. Correlation:

$$\frac{\sum_{a,b}[(ab)P_{\phi,d}(a,b)] - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{13}$$

where:

$$\mu_x = \sum_a a \sum_b P_{\phi,d}(a,b) \tag{13a}$$

$$\mu_y = \sum_b b \sum_b P_{\phi,d}(a,b) \tag{13b}$$

$$\sigma_x = \sum_a (a-\mu_x)^2 \sum_b P_{\phi,d}(a,b) \tag{13c}$$

$$\sigma_y = \sum_b (b-\mu_y)^2 \sum_b P_{\phi,d}(a,b) \tag{13d}$$

A feature vector of skin image based on GLCM is 32-dimension has been obtained from eight features and each of features constructed by four angles directions.

2.4. Distance Metric and LVQ Neural Network Classifier

A feature vector $F = \{f_1, f_2, ..., f_n\}$ is mapped to a point $P(f_1, f_2, ..., f_n)$ in the n-dimensions. This mapping helps us to perceive the images (represented by their feature vectors) as high-dimensional points. The advantage of this representation is that one can now use different distance metrics for (i) finding similarity between two images and (ii) ordering a set of images based on their distances from a given image. This research considers Manhattan distance and Euclidean distance [10] and also Learning Vector Quantization (LVQ) Neural Network as a classifier [11].

Manhattan distance is also called the L_1 distance. If $u = (x_1, y_1)$ and $v = (x_2, y_2)$ are two points, then the Manhattan distance between *u* and *v* is given by

$$d_M(u,v) = |x_1 - x_2| + |y_1 - y_2| \tag{14}$$

Instead of two dimensions, if the point has n-dimensions, such as $a = (x_1, x_2, ..., x_n)$ and $b = (y_1, y_2, ..., y_n)$ then the Manhattan distance between *a* and *b* could be written as

$$d_M(a,b) = |x_1 - y_1| + |x_2 - y_2| + ... + |x_n - y_n|$$
$$= \sum_{i=1}^{n} |x_i, y_i| \quad (15)$$

Euclidean distance is also called the L_2 distance. If $u = (x_1, y_1)$ and $v = (x_2, y_2)$ are two points, then the Euclidean distance between *u* and *v* is given by

$$d_E(u,v) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (16)$$

Instead of two dimensions, if the point has n-dimensions, such as $a = (x_1, x_2, ..., x_n)$ and $b = (y_1, y_2, ..., y_n)$ then the Euclidean distance between *a* and *b* could be written as

$$d_E(a,b) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + ... + (x_n - x_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (17)$$

An LVQ neural network is a supervised version for a training competitive layer for adaptive pattern classification. A competitive layer will automatically learn to classify input vectors. However, the classes that the competitive layer found are dependent only on the distance among the input vectors. If two input vectors are very similar, the competitive layer will probably put them into same class. The tuning of the decision surface is done by rewarding the correct classifications and punishing the incorrect ones. When training pattern x^k from class c_j is presented to the network, let the closest reconstruction w_i belongs to class c_l. Then only vector w_i is updated according to the following supervised rule,

$$\Delta w_i = \begin{cases} +\eta^k (x^k - w_i) & if \quad c_j = c_l \\ -\eta^k (x^k - w_i) & if \quad c_j \neq c_l \end{cases} \quad (18)$$

where the learning rate η^k is assumed to be monotonically decreasing function of the number of iterations *k*. The decreasing learning rate allows the network to converge the network to a state in which the weight vectors are stable. The primary effect of equation (18) is to minimize the number of misclassifications. At the same time, the vectors w_i are pulled away from the zones of class overlap where misclassifications persist. The algorithm described here is referred to LVQ1 learning rule. The LVQ network architecture is shown in Figure 4.

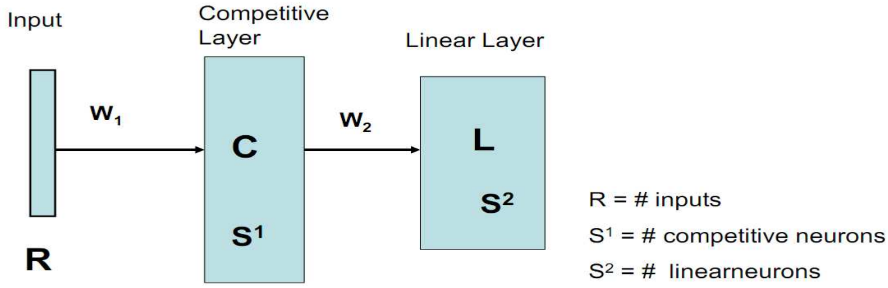

Figure 4. The LVQ network architecture.

3. Results and Analysis

The purpose of the feature extraction is to enhance the variability which helps to discriminate among the classes. It means the features vector is found to be very different between skin image non- irritation and irritation. The gray level histogram (GLH) features have been selected and the feature vector of image skin obtained is shown in Figure 5.

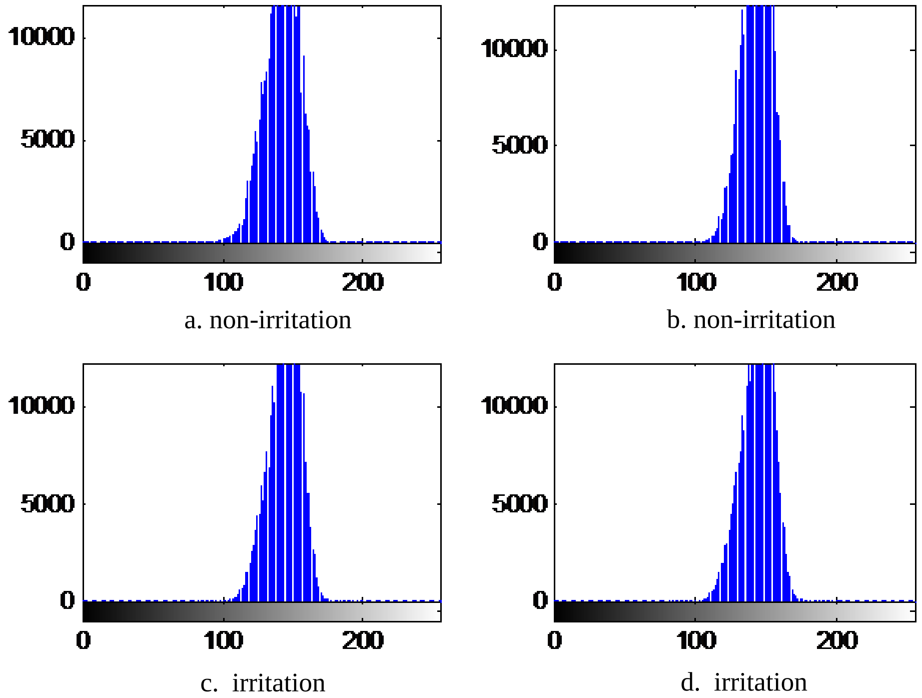

Figure 5. Gray Level Histogram of skin images

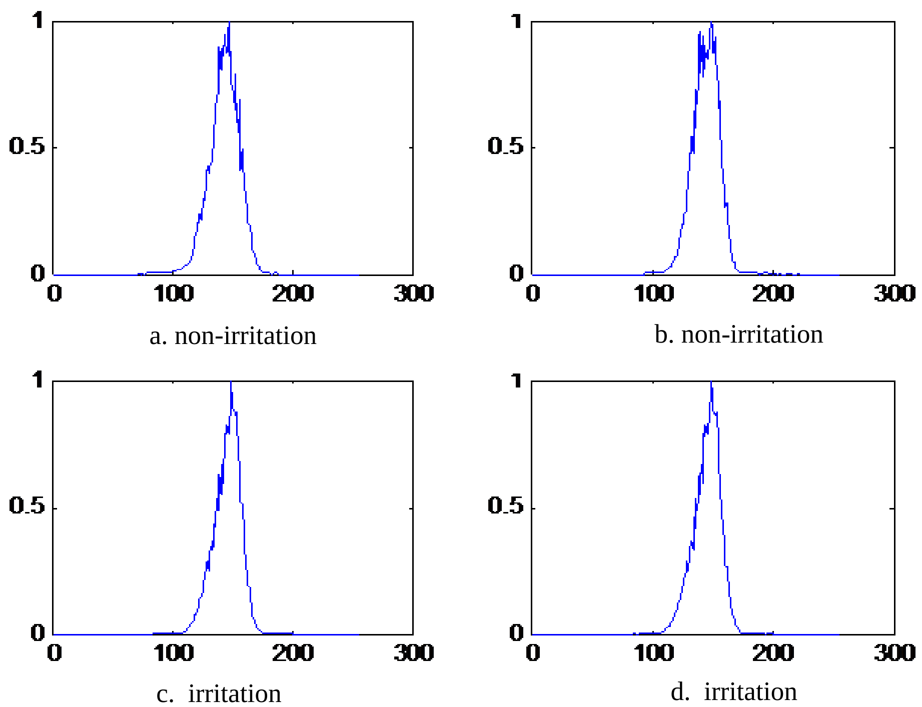

Figure 6. Gray Level Histogram of skin images

For enhancement the features vector, it have been proposed new approach namely a normalized the feature vector. The normalization of the GLH fatures is obtained by dividing the the maximum value of histogram as shown in Figure 6. The GLCM features are shown in Figure 7.

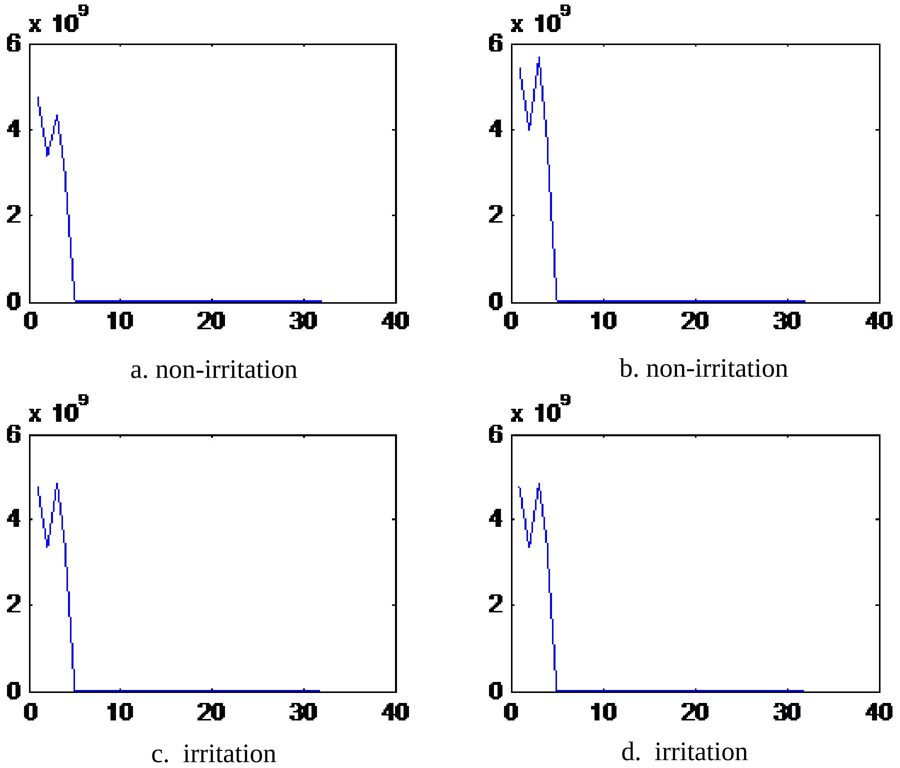

Figure 7. Gray Level Co-occurrence Matrix of skin images

Table 2. The experimental results of combining features and classifier

Features	Accuracy (%)		
	Distance Metric		LVQ neural network
	Manhattan	Euclidean	
Gray Level Histogram	62.50	50	83.33
Gray Level Co-occurance Matrices	75	75	83.33

The methods for identification of skin irritation have been implemented in MATLAB programming language. The data collection have considered different human skin images between non-irritation or irritation. The all database that it's used in this experiment as shown in Table 1.

Table 3. The experimental results of GLH and LVQ neural network

Learning rate (α)	**0,1**	**0,3**	**0,5**	**0,7**	**0,9**
# Competitive neurons = 32					
Accuracy (%)	79,2	66,7	79,2	50	50
# Competitive neurons = 64					
Accuracy (%)	70,8	**83,3**	66,7	50	50
# Competitive neurons = 128					
Accuracy(%)	70	58,3	70	66,7	50
# Competitive neurons = 256					
Accuracy (%)	58,3	58,3	75	45	50

Table 4. The experimental results of GLH and LVQ neural network

Learning rate (α)	**0,1**	**0,3**	**0,5**	**0,7**	**0,9**
# Competitive neurons = 16					
Accuracy (%)	75	66.67	70.83	62.5	50
# Competitive neurons = 32					
Accuracy (%)	70.83	**83.33**	70.83	66.67	58.33
# Competitive neurons = 64					
Accuracy (%)	75	75	66.67	75	62.5

All of the experiments were performed using combination of features and classifier methods. The experimental results are shown in Table 2. From that table, the best accuracy of 83.33% is obtained when the design system is implemented for GLH or GLCM features using an LVQ neural network classifier.

The performance system is found the be best accuracy when the identification system design uses LVQ neural network, but it's not simple during the training phase compare than the distance metric. In the training phase of the LVQ neural network done by trial and error methods by adjusting the learning rate α or training cycles. This strategy will enable to obtain more accurate results. And also, the number of competitive neurons have been adjusted to obtain the optimal of accuracy. The detail results of the experiment based on an LVQ neural network classifier are shown in Table 3 and Table 4.

4. Conclusion

In this paper, an automatic human skin identification system based on an image recognition has been proposed. The system was objectively developed to be cost-effective to replace skin irritation identification manually. The performance system has been evaluated using a combination of features and classifier methods. In the experiment, the design have been applied using the GLH and GLCM features extractor. The selected classifiers were a distance metric (Manhattan and Euclidean),and LVQ neural network. From that selection, the system shows a high rate of accuracy for 83.33% through LVQ neural network classifier disregard using GLH or GLCM feature. In the future, it is still possible to improve its accuracy using new feature extractor techniques and classifier methods.

References

[1] Hashim P, Shahab N, Masilamani T, Baharom R, Ibrabim R. A Cosmetic Analysis in Compliance with the Legislative Requirements, Halal and Quality Control. *Malaysian Journal of Chemistry*. 2009; 11(1): 081-087.

[2] Alsmadi MKS, Omar KB, Noah SA, Almarashdeh I. Fish Recognition Based on Robust Features Extraction from Size and Shape Measurements Using Neural Network. *Journal of Computer Science*. 2010; 6(10): 1059-1065.

[3] Khalid M, Lee ELY, Yusof R, Nadaraj M. Design of An Intelligent Wood Species Recognition System. *International Journal Simulation System, Science & Technology*. 2008; 9(3): 9-19.

[4] Mostafa L, Abdelazeem S. *Face Detection Based on Skin Color using Neural Networks*. Preoceeding of GVIP. Cairo, Egypt. 2005: 51-56.

[5] Umarani C, Ganesan L, Radhakrishnan S. Combineed Statistical and Structural Approach for Unsupervised Texture Classification. *International Journal of Imaging and Engineering*. 2008: 2(1): 162-165.

[6] Cula OG, Dana KJ, Murphy FP, Rao BK. Skin Texture Modeling. *International Journal of Computer Vision*. 2005; 62(1/2): 97-119.

[7] Haralick RM. Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 1979; 67(5): 786-804.

[8] Popescu D, Dobrescu R, Nicolae M. Texture Classification and Defect Detection by Statistical Features. *International Journal of Circuits, system and Signal Processing*. 2007; 1(1): 79-84.

[9] Gonzalez RC, Woods RE. Digital Image Processing. Second edition. New York: Prentice Hall. 2001.

[10] Fadlil A. Simple Program Face Recogniton System Using Distance Function. *TELKOMNIKA*. 2006; 4(3): 153-158.
[11] Demuth H, Beale M. Neural Network Toolbox. New York: The Math Works Inc. 1994.