Journal of Advances in Mathematics and Computer Science

25(6): 1-10, 2017; Article no.JAMCS.29634

ISSN: 2456-9968

(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Iris Texture Analysis for Ethnicity Classification Using Self-Organizing Feature Maps

B. M. Latinwo[1], A. S. Falohun[1*], E. O. Omidiora[1*] and B. O. Makinde[2]

[1]Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.

[2]Department of Computer Science, Osun State College of Technology, Esa-Oke, Osun State, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMCS/2017/29634

Editor(s):

(1) Doina Bein, Professor, Applied Research Laboratory, The Pennsylvania State University, USA.

Reviewers:

(1) S. K. Srivatsa, Anna University, India.

(2) Mohamed Zorkany, National Telecommunication Institute, Egypt.

(3) Remus Brad, Lucian Blaga Univeristy of Sibiu, Romania.

(4) Chelali Fatma Zohra, University of Science and Technology, Algeria.

Complete Peer review History: http://www.sciencedomain.org/review-history/22581

Original Research Article

Received: 21st September 2016
Accepted: 5th November 2016
Published: 3rd January 2018

Abstract

Ethnicity Classification from iris texture is a notable research in the field of pattern recognition that differentiates groups of people as distinct community by certain characteristics and attributes. Several ethnicity classification systems have been developed using Supervised Artificial Neural Network and Machine Learning algorithms. However, these systems are limited in their clustering ability and require prior definition of image classes which lowers its classification rate. Therefore, this work classified iris images from Nigeria, China and Hong Kong origin using Self-Organizing Feature Maps (SOFM) blended with Principal Component Analysis (PCA) based Feature extraction. Left and right irises of 240 subjects constituting 480 images were acquired online from CUIRIS (Nigeria), CASIA (China) and CUHK (Hong Kong) datasets, and normalized to a uniform size of 250 by 250 pixels. Three hundred and thirty six (336) images were used for training while the remaining 144 were used for testing. The system was implemented in Matrix Laboratory 8.1 (R2013a). The performance of the classification system was evaluated at varying thresholds (0.2, 0.4, 0.6 and 0.8) and 93.75% Correct Classification Rate (CCR) was obtained.

**Corresponding author: E-mail: asfalohun@lautech.edu.ng, eoomidiora@lautech.edu.ng;*

Keywords: Ethnicity; iris-texture; image classification; SOFM.

1 Introduction

The iris is the coloured portion of the eye visible around the pupil and it is covered by the cornea. And it is the only internally protected organ of the body that is visible from the outside. It is formed before birth via a process known as chaotic morphogenesis and is very stable from birth to death. Iris texture pattern can be used for biometric verification and identification of a person from a large dataset. However, through investigation of large number of iris images of different races, [1] found that iris patterns have different characteristics on the overall statistical measurement of the iris texture.

Whereas ethnicity is defined as the fact or state of belonging to a social group that has common national or cultural tradition. But sometimes the definition for race is somewhat equated with ethnicity; as human race can be defined as a group of people with certain inherited features that distinguish them from other groups of people. All men of whatever race are classified by the anthropologist or biologist as belonging to one specie, homosapiens. Ethnicity can also be defined as a vast group of people loosely bounded together by historical, socially significant elements of their morphology and/or ancestry. It can serve as the connections between physical features, races and personal characteristics [2]. According to [3], there are 4 ethnographic division of man to races in existence today; Caucasian (Whites), Mongoloid (Asians), Negroid (blacks) and Australoid.

In the area of classification, Image classification is an important and challenging task in various application domains, including biomedical imaging, biometry, video surveillance, vehicle navigation, industrial visual inspection, robot navigation, and remote sensing [4]. Ethnicity classification from iris texture is also an active and expanding research area that has made use of these several image classification techniques [1,5] and [6,7,8,9]. It is an old topic in social science and often assumed to be a fixed trait based on ancestry. In natural science, few attempts have been made to perform automatic ethnicity classification based on human images. But in Computer Science, ethnicity classification makes use of different algorithms and it includes three basic modules: image preprocessing, feature extraction, and training [1].

Image classification techniques applied in the area of ethnicity classification from iris texture have been majorly the supervised techniques of Artificial Neural Network (Support Vector Machine, Adaboost learning algorithm, Multilayer perceptron, etc.), Weka machine learning algorithms (Naives Bayes, J48, and the likes) [8] and Principal Component Analysis [10]. These techniques had inadequate graphical representation for clustering, failed to produce meaningful clusters, are sensitive to initialization techniques and require prior definition of image classes. Amongst the unsupervised aspect of the ANN algorithm, K-means has been applied. These various techniques were applied on Asian and Caucasian irises. SOFM technique has not been applied in the area of ethnicity classification from iris texture despite being mostly used in pattern matching in iris and facial recognition systems. Hence, this work explored an iris-involved ethnicity classification through SOFM.

2 Literature Review

2.1 Self- organizing feature map

[11] proposed and provided Self Organizing Feature Maps (SOFM) as an orderly mapping of an input high-dimensional space in much lower dimensional spaces, usually one or two dimensions. It compresses information while preserving the most important topological and metric relationships of the primary data; it produces some kind of abstractions of information. It can also be utilized in a number of ways in complex tasks such as pattern classification, process analysis, machine perception, control and communication.

SOFMs are different from other ANNs in the sense that it uses a neighborhood function to preserve the topographical properties of the input space. This makes SOFM useful for visualizing low dimensional views of high-dimensional data. Like most ANN, SOFM operate in two modes: training and mapping. Training builds the map using input example. It is a competitive process also called vector quantization. Mapping automatically classifies a new input vector. A self-organizing feature map consists of components called nodes and neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in hexagonal or rectangular grid. The SOFM describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from data space and to assign the map coordinate of the node vector [12].

The self-organization process involves four major components:

1. Initialization: All the connection weights are initialized with small random values.
2. Competition: For each input pattern, the neurons compute their respective values of a discriminant function which provides the basis for competition. The particular neuron with the smallest value of the discriminant function is declared the winner.
3. Cooperation: The winning neuron determines the spatial location of a topological neighborhood of excited neurons, thereby providing the basis for cooperation among neighboring neurons.
4. Adaptation: The excited neurons decrease their individual values of the discriminant function in relation to the input pattern through suitable adjustment of the associated connection weights, such that the response of the winning neuron to the subsequent application of a similar input pattern is enhanced [13]. A typical architecture of the technique is shown in Fig. 1.

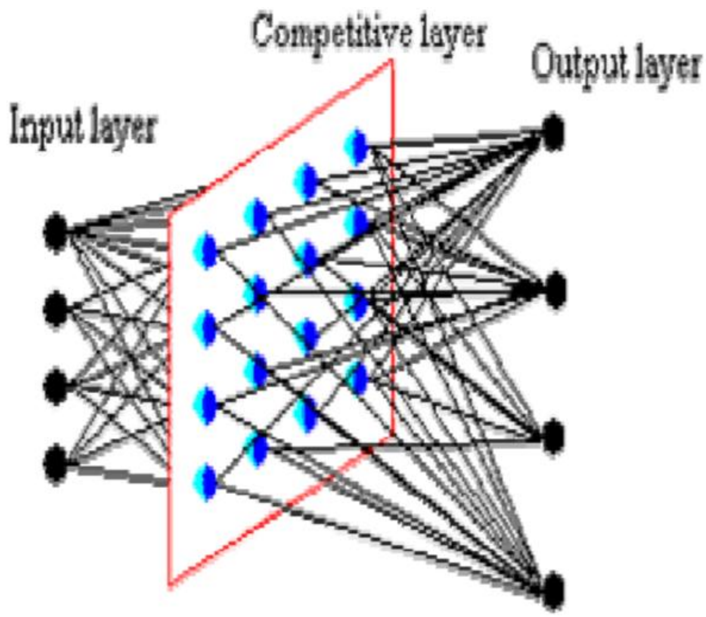

Fig. 1. Architecture of SOFM [14]

2.2 Review of related works

Quite a number of researches are seen in this area of endeavor some of which are as described in Table 1.

Table 1. Classification techniques that have been used for ethnicity prediction from iris texture

Author(s)	Classification technique used	Iris images used and the result of the research	Classification accuracy
Qui, Sun and Tan (2006)	Adaboost Learning Algorithm	600 randomly selected images from Asian and Non-Asian respectively making 1200 were used for training set and 2782 images for test set and were classified using CCR and established that the global texture features of iris are efficient for ethnic classification.	85.95%
Qui, Sun and Tan (2007b)	Support Vector Machine (SVM)	2400 images of 60 persons from Asian/Caucasian ethnicity and used 1,200 images as training set and 1,200-image test set). The work established that if iris images from same person appear in the training and test, the performance estimate obtained is optimistically biased.	91%
Qui, Sun and Tan (2007b)	Texton features (K-means Algorithm)	400 iris images of Asian and Non-Asian for learning	95%
Stark, Bowyer and Siena (2010).	ANN perceptual categorization	100 iris images of 100 persons from Asian/Caucasian ethnicity and established that human observers perceive consistent ethnicity-related difference in iris texture.	80%
Lagree and Bowyer (2011)	All algorithms of WEKA packages were tried (SMO, Bagged FT, J48, Naive Bayes) SMO gave the highest accuracy and statistics	1200 iris of the left and right irises from 60 persons of Caucasian and/Asian ethnicity. A 1-N recognition application was developed to match a prove iris code against a large number of enrolled identities for recognition, prediction of demographic factors for search ordering and thereby reducing average search. SMO gave the highest accuracy and statistics.	90.58%

3 Methodology

The basic stages utilized in this research work are: Acquisition of iris images, selection of training and testing data, preprocessing the images, feature extraction, classification and performance evaluation in 250x250 pixel dimensions at 0.2, 0.4, 0.6 and 0.8 thresholds.

3.1 Image acquisition

Iris images from three sources were used for the development of this system, the samples are from three datasets: CUIRIS version 1 (African), CASIA version 1 (Asia) and CUHK version 1 (Asia), these images are in bitmap (.bmp) format [15–17]. The images were selected at random, 144 testing and 336 training images of 80 subjects (left and right irises) from the three ethnicities was used for ethnicity classification.

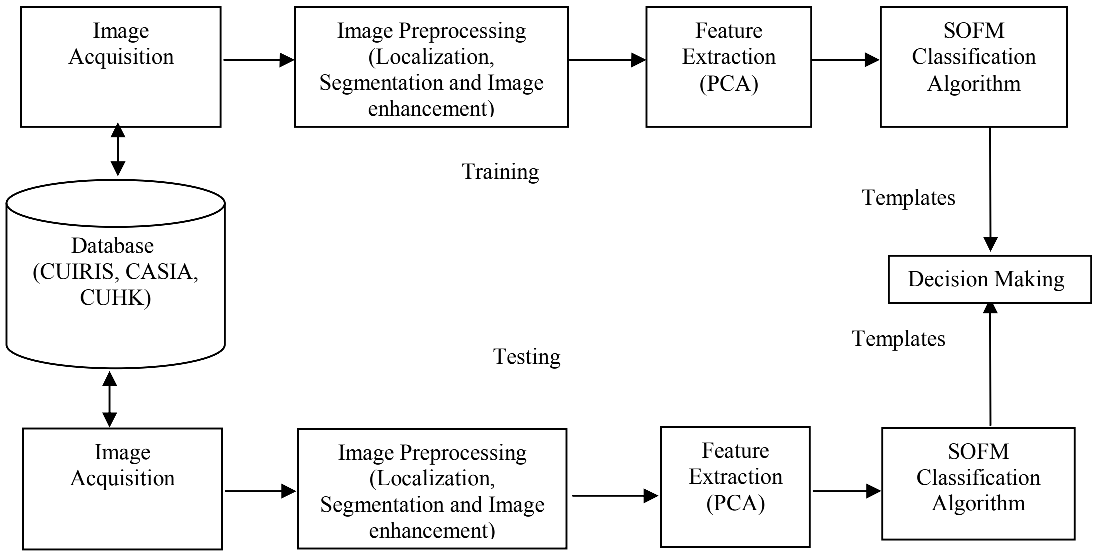

Fig. 2. Process flow diagram for the developed ethnicity classification system

3.2 Image preprocessing

The Image preprocessing module include: iris image localization, segmentation and enhancement. Localization involves locating the iris in an eye image while segmentation involves detection and exclusion of occluding eyelids, eyelashes or reflections to isolate the actual iris from the eye region. It is also the process of decomposing the images into regions and objects by associating or labeling each pixel with the object that it corresponds to. Circular Hough Transform approach was used for the iris segmentation to deduce the radius and centre coordinates of the pupil and iris regions.

Due to illumination variations, the radial size of the pupil usually change accordingly. The resulting deformations of the iris texture always affect the performance of the subsequent, feature extraction and matching stages. Therefore, the iris region was enhanced to compensate for variations using Histogram Equalization thus improving the images by noise removal and contrast enhancement.

3.3 Image segmentation using circular Hough transform

Hough Transform Segmentation technique was used to isolate the iris from the eye region since [18] reported its better performance than other localization techniques in case of occlusion due to eyelids and eyelashes. It is an automatic segmentation algorithm employed to deduce the radius and center coordinates of the pupil and iris region. It’s parameters are the centre coordinates xc and yc, and the radius r, which are able to define any circle according to the equation

$$x_c^2 + y_c^2 - r^2 = 0 \tag{1}$$

It also detect eyelids, approximating the upper and lower eyelids with parabolic arcs, which are represented as (2)

$$(-(x-h_j)\sin\theta_j + (y-kj)\cos\theta_j)^2 = a_j((x-h_j)\cos\theta_j + (y-kj)\sin\theta_j \tag{2}$$

Where aj controls the curvature, (hj, kj) is the peak of the parabola and Ø is the angle of rotation relative to the x-axis.

3.4 Feature extraction using PCA

Feature dimensionality reduction and extraction of unique features of the segmented and enhanced images were done using PCA, a multivariate analysis technique used for reduction of large dimension numbers before subjecting output factors to a clustering routine. Eigenvectors used are facts that the principal components of the training set of iris images generated after reducing the dimensionality of the training set. Once they are selected, each training set image is represented in terms of these eigenvectors. Then, an unknown iris, used for recognition purposes is represented in terms of the eigenvectors. The eigenvectors representation of this unknown iris is compared with each training set iris image. The "distance" between them is then calculated. If the distance is above a specified threshold value, the unknown iris is recognized as that person and vice versa.

3.5 Training and testing with SOFM

Self-Organizing Feature Maps (SOFM) formed its own classifications of the training and testing data without external help (being an unsupervised Neural Network model). During the training phase, the input vectors were presented based on the initial weights chosen at random, the neuron with weights closest to the input image vector was declared as the winner. Then the weights of all the neurons in the neighborhood of the winning neuron was adjusted by an amount inversely proportional to the Euclidean distance, that is, the distance between the input image vector and the weight vector. It eventually clustered and classified the vectors based on the attributes used. Its components are: initialization, competition, cooperation and adaptation discussed in section 2.

The steps in the SOFM algorithm are as follows:

Step 1: Topological neighborhood parameters, learning rate, initialize weights was set.
Step 2: While stopping condition is false, steps 3 to 9 was done.
Step 3: For each input vector x, steps 4 to 6 was done.
Step 4: For each j, squared Euclidean distance was computed.
$D(j) = \ (wij - xi)^2$, i = 1 to n and j = 1 to m.
Step 5: Index j was found, when D (j) is minimum.
Step 6: For all units j, with the specified neighborhood of j, and for all i, the weights were updated.
Step 7: Learning rate was updated.
Step 8: Radius of topological neighborhood was reduced at specified times.
Step 9: The stopping condition was tested.

From the database of preprocessed iris images, SOFM trained and tested the iris images and organized them based on content. Utilizing its intrinsic properties, related irises based on the threshold range set were found close to each other in the network.

The similarity based classifier found the best matching (winning) neuron at a time in a self-organizing feature map by using the minimum distance Euclidean criterion with thresholding in its unsupervised state to map iris images from the same ethnicity together. During this matching and classification phase, each input was compared with all nodes of the SOFM, and the best match was found and the final output of the system based on similar iris codes was displayed. And then the correctly classified images of each ethnicity were calculated based on the result given by the SOFM tools in a MATLAB environment.

3.6 Performance evaluation

The performance of the ethnicity classification system developed was measured using True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN). Which were used to calculate false positive rate, false negative rate, positive prediction value, negative prediction value and correct classification rate.

$$\text{Sensitivity: } \frac{TP}{(TP+FN)} \tag{1}$$

$$\text{Specificity: } \frac{TN}{(TN+FP)} \tag{2}$$

$$\text{Positive Predictive Value: } \frac{TP}{(TP+FP)} \tag{3}$$

$$\text{Negative Predictive Value: } \frac{TN}{(TN+FN)} \tag{4}$$

$$\text{Correct Classification Rate: } \frac{(TP+TN)}{(TP+TN+FN+FP)} \tag{5}$$

Correct Classification Rate (CCR) is the number of correctly classified claims divided by the total number of claims in the dataset. Sensitivity is the probability that the test indicated the presence of an iris image belonging to an ethnicity in a created database. Specificity is the probability that the test indicated the presence of an iris image belonging to an ethnicity but tested negative for that ethnicity in the created database. PPV and NPV are influenced by the prevalence of an ethnicity in the population being tested.

3.7 Implementation in MATLAB

The implementation of ethnicity classification approach follows series of operations. At the point of execution, the GUI window popped up and the images to be analyzed were selected. The interface was made explicit with the help of some created buttons such as “Train button”, “Classify button”, and so on as shown in Fig. 3. In Figs. 3 and 4, the aftermath of clicking the “Train button” by selecting training with SOFM was seen. Clicking the “Train button” performs the loading operation, pre-processing operations; feature extraction process concurrently and finally stored the executed results. The “Test button” on the click, also performed preprocessing of its images and stored their templates and then classified individual images based on correctly or incorrectly identified image using confusion metrics such as Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value and finally determined the correct classification rate coupled with the computation time. These buttons operate based on the SOFM algorithm.

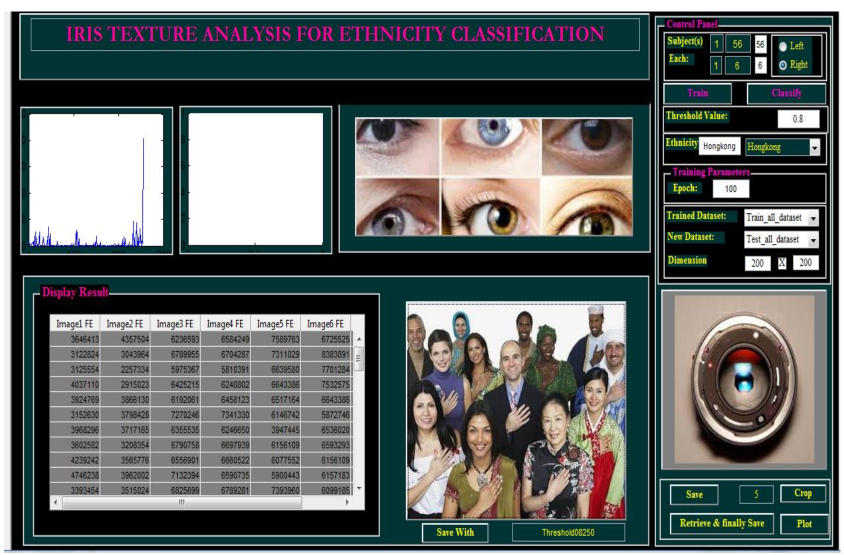

Fig. 3. MatLab GUI showing the training stage of iris images at different thresholds and pixel dimensions

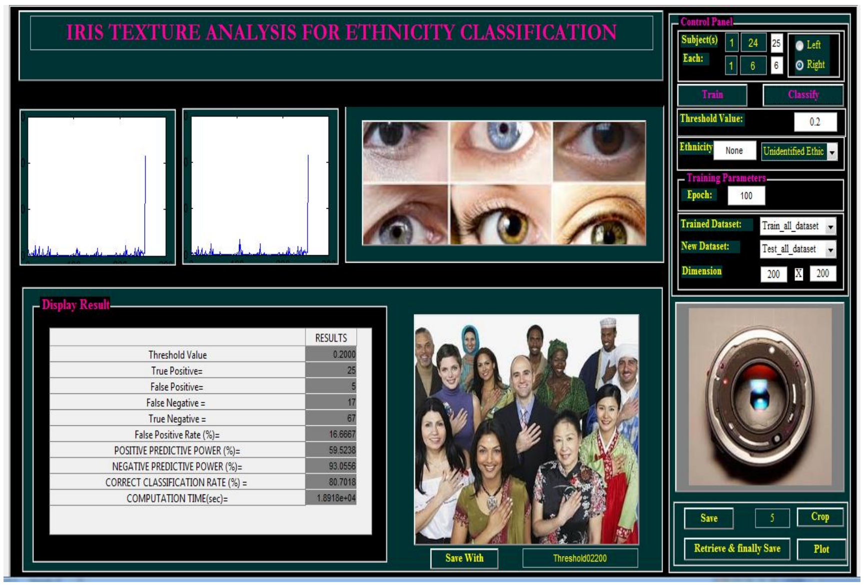

Fig. 4. MatLab GUI showing the testing and classification process of iris images at different thresholds and pixel dimensions

4 Results and Discussion

Results for the testing and classifications by SOFM generated values for true positive, true negative, false positive and false negative which were used to calculate correct classification rate based on the standard formula stated in equation (5). The CCR result at thresholds 0.2, 0.4, 0.6 and 0.8 in 250 by 250 pixels as shown in Table 2 are 86.81, 90.27, 93.06, and 93.75% respectively. These values were based on the similarity mapping by SOFM.

Table 2. Testing and classification result

Threshold	TP	FP	FN	TN	CCR (%)	Time/s
0.2	56	4	15	69	86.81	36.92
0.4	57	3	11	73	90.27	38.10
0.6	58	2	8	76	93.06	40.20
0.8	58	2	7	77	93.75	42.62

4.1 Discussion

The classification phase where SOFM mapped similar images together as seen from Fig. 5 showed the topological and clustering effect of this technique where similar test images were clustered together based on characteristics inherent in them. The layer of the map showed test data as grey-blue patches and joined similar ones with red lines, the other dissimilar test data are colored from black to yellow. It partitioned these test images into a number of classes based on statistical information inherent to them as seen in Table 2. This is the basis for classification as true positive, true negative, false positive and false negative, in which their values varied depending on whether images belonging to an ethnicity were effectively matched and vice-versa. There was also possibility of images not belonging to an ethnicity being matched to it because of their common ethnic region.

The values generated for the TP, FP, FN and TN was used to compute the CCR at the four thresholds where the highest performance evaluation result was obtained at 0.8. This established that the higher the threshold, the more the iris texture features can be visible for analysis and effective classification.

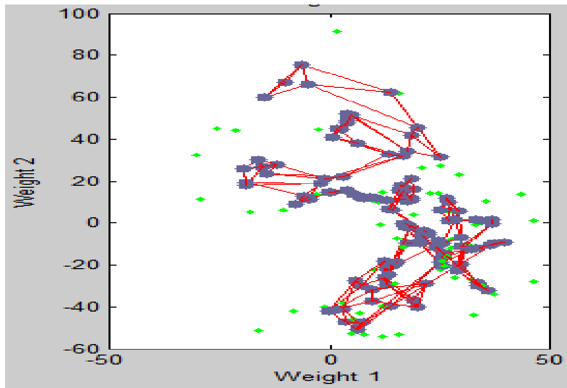

Fig. 5. Diagram showing the mapping of similar test images

5 Conclusion

In this research work, iris texture analysis for ethnicity classification using Self-Organizing Feature Maps was carried out. The training and testing images were preprocessed using Hough transform, Histogram Equalization while PCA extracted the principal components and SOFM classified the test images yielding a classification accuracy of 93.75% at the highest chosen threshold. This compares favorably with earlier ethnicity classification works , especially that of Qui, Sun & Tan (2007b) where an accuracy of 95% was obtained using K-Means Algorithm with 400 Asian and Non- Asian iris images. The difference might be because of the presence of CUIRIS, the African dark-eye set with characteristic poor contrast.

6 Recommendation

Although this research was not performed on a very large scale of data, it has established the strength of SOFM. More of the strengths and weaknesses of SOFM algorithms can be investigated by using it on a very large dataset with multiple ethnicities. It can also be used for possible prediction of other demographic factors such as age and gender. Also, other unsupervised ANN algorithms can be tried for ethnicity classification to ascertain the likely outcome.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Qiu XC, Sun ZA, Tan TN. Global texture analysis of iris images for ethnic classification. Springer LNCS 3832: International Conference on Biometrics. 2006;411-418.

[2] Ratcliffe P. Ethnic group: ISA Editorial Arrangement of sociopedia.isa.
DOI: 10.1177/205684601421.2014

[3] McClure S, Poole M, Anderson-Fye E. Anthropological approaches to the study of body image and human appearance. In Encyclopedia of Body Image and Human Appearance, Eds., T. Cash, et al. Elsevier Press; 2012.

[4] Kamavisdar P, Saluja S, Agrawal S. A survey on image classification approaches and techniques. International Journal of Advanced Research in Computer and Communication Engineering. 2013;2:1.

[5] Qiu XC, Sun ZA, Tan TN. Learning appearance primitives of iris images for ethnic classification. International Conference on Image Processing. 2007a;II:405–408.

[6] Qiu XC, Sun ZA, Tan TN. Coarse iris classification by learned visual dictionary. Springer LNCS 4642, International Conference on Biometrics. 2007b;770–779.

[7] Stark L, Bowyer KW, Siena S. Human perceptual categorization of iris texture patterns. Biometrics: Theory, Applications and Systems, Washington, DC: BTAS; 2010.

[8] Lagree S. and Bowyer K.W. Ethnicity Prediction Based on Iris Texture feature. Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556 USA. 2011

[9] Zarei A, Duxing M. ANN for prediction of ethnicity based on iris texture. Department of Electrical Engineering, University of the Pacific Stockton, CA, USA; 2012.

[10] Latinwo BM, Falohun AS, Omidiora EO. Principal component analysis - based ethnicity prediction using iris feature. British Journal of Applied Science & Technology. 2016;16(4):1-5. SCIENCEDOMAIN International. 2016.

[11] Kohonen T. Self-organizing maps. Springer Series in Information Sciences. 1995;30:974–978.

[12] Kohonen T, Honkela T. Kohonen Network. Scholarpedia; 2007.

[13] Bullinaria JA. Self-organizing maps: Fundamentals. Introduction to Neural Networks; 2004.

[14] Savita S, Sharda V, Asha G, Anjali G. Iris pattern recognition using self- organizing neural network. International Journal of Computer Applications; 2012.
ISSN: 0975-8887.

[15] CASIA: Chinese Academy of Sciences Institute of Automation (CASIA) Iris Database; 2005.
Available:http://www.cbsr.ia.ac.cn/english/irisDatabase.asp

[16] CUHK: Computer Vision Laboratory, Department of Automation & Computer-Aided Engr., The Chinese University of Hong Kong (2003), Hong Kong.

[17] CUIRIS: Covenant University Iris (CUIRIS) iris database; 2012.
Available:http://drive.google.com/CUirisdatabase.zip

[18] Dolly Choudhary, Shamik Tiwari, Ajay Kumar Singh. Survey: Feature extraction methods for iris recognition. International Journal of Electronics Communication and Computer Technology (IJECCT). 2012;2(6):64-72.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/22581