Principal Component Analysis - Based Ethnicity Prediction Using Iris Feature

B. M. Latinwo¹, A. S. Falohun¹* and E. O. Omidiora¹*

¹Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, 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/BJAST/2016/26131

Editor(s):
(1) Samir Kumar Bandyopadhyay, Department of Computer Science and Engineering, University of Calcutta, India.

Reviewers:
(1) Irena Ilieva Jekova, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Bulgaria.
(2) R. Gayathri, Sri Venkateswara College of Engineering, Anna University, Sriperumbudur, India.

Complete Peer review History: http://sciencedomain.org/review-history/14965

Received 1st April 2016
Accepted 1st June 2016
Published 9th June 2016

ABSTRACT

This paper presents the effectiveness of Principal Component Analysis (PCA) technique in analyzing iris texture by performing dimensionality reduction and extracting unique feature codes of images for efficient ethnicity classification using iris images from African and two Asian datasets. Three hundred and thirty-six iris images were obtained, preprocessed (enhanced) and segmented for easy identification of unique features using Histogram Equalization and Hough Transform techniques, respectively. Feature dimensionality reduction and extraction of feature codes of the segmented images was carried out using PCA while the result showed the similarities and differences between irises of different ethnicities based on these generated codes. The research established a very close similarity of the Asia1 and Asia2 irises, due to the classification of their features in the same feature code subrange. Also, few images from Asia1 and Asia2 were classified with the Africans which explained the possibility of mixed race of subjects through inter-marriage.

Keywords: Ethnicity; Iris; feature codes; PCA; classification.
1. INTRODUCTION

Among the various biometric technologies (fingerprint, iris, face, palm print, hand geometry, gait, etc.), iris recognition system has been acclaimed to be highly accurate, reliable and fool-proof because irises are highly distinctive and of stable characteristics throughout lifetime [1, 2] except in the case of cataract surgery, accidents or other eye-related problems [3]. Just like fingerprints, irises are unique to each individual and have little similarities between ethnic groups [4] and worthy of mention is the fact that researchers have been working from the last decade to extend the application of iris recognition system to several areas like tracing criminals, terrorist and missing children; ethnicity, age and gender prediction; accurate diagnosis of eye defect and ascertaining state of health [5]. A notable access control research based on biometric traits that combined black and white iris images in an online mode (page 204 in [6]) enabled remote accessing with iris images while a deployment of biometrics to control engineering was equally explored in using iris signature to unlock a door system (page 522 in [7]).

Ethnicity or ethnic group is a socially defined category of people who identify with each other based on common ancestral, social, cultural or national experience. They are groups of people that regard themselves or are regarded by others as a distinct community by virtue of certain characteristics that helps to distinguish them from the surrounding community. It is considered to share characteristics such as culture, language, religion, and traditions which contribute to a person or group’s identity. Ethnicity is also a preferential term to describe the difference between humans rather than race [8, 9], exposed 4 ethnographic division of man to races in existence today; Caucasian races (Whites), Mongoloid races (Asians), Negroid races (blacks) and Australoid. Ethnicity classification from iris texture is also an active and expanding research area that has made use of several image classification techniques such as [10-13]. The authors utilized Asian and Caucasian iris images with image classification algorithms whereas present work analyzed and predicted the ethnic groups only at feature extraction stage through feature codes.

Badejo [14] identified the peculiarity of dark-brown iris images of Africans in comparison with Asians and Caucasians using histogram and normal probability distribution of their grayscale image entropy (GiE) values among other works whereas iris or feature codes have not been explicitly analyzed for more clarification for the establishment of their distinct differences from other ethnic origins apart from eye color discrimination. These significant features of the iris texture pattern must be encoded so that comparisons between templates can be made conveniently and correctly [15]. Here, PCA was employed for the feature code extraction because of its ability to emphasize variation and bring out strong patterns in a dataset, thus making data easy to explore and visualize. Therefore, this research attempted a classification of iris images from African and Asian origins using Principal Component Analysis.

2. RESEARCH METHODOLOGY

In this paper, the three stages of the development of the ethnicity prediction system are: Image acquisition, preprocessing and feature extraction. Iris images were acquired from CUIRIS, CASIA and CUHK datasets [16-18] which later got preprocessed. Hough Transform was used for segmentation, Histogram Equalization for the image enhancement and PCA extracted the important features. The performance was tested based on the feature codes of these ethnicities. The result showed a distinct difference between African (CUIRIS) and Asian irises and the similarities between Asia1 (CASIA) and Asia2 (CUHK) irises.

2.1 Image Acquisition

Iris images from three sources were used for the development of this system. The samples were from CUIRIS, CASIA and CUHK which represented Africa, Asia1 and Asia2 ethnicity respectively. Left and right iris images of 168 subjects (2 images per person) were used from the three datasets, making 336 images in all. They were used in 250x250 pixel dimensions.

2.2 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. 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 Hough Transform approach was used to deduce the radius and center coordinates of the pupil and iris regions.

Due to illumination variations, the radial size of the pupil will change accordingly. The resulting deformation of the iris texture will affect the performance of subsequent, feature extraction stage. Therefore, Image enhancement was performed to compensate for the inevitable variations using Histogram Equalization.

2.3 Feature Extraction

Feature extraction is a special form of dimensionality reduction which contains more information about the original image. Features are extracted using preprocessed iris images and the most discriminating information with the PCA equations.

2.3.1 Principal Component Analysis (PCA)

It is a multivariate analysis technique used for reduction of large dimension numbers before subjecting output factors to a clustering routine. It is used to extract size-dimension information and to construct a linear projection of a dataset and for processing and visualizing data [19,20]. It extracts the main variation in the feature vector and allows an accurate reconstruction of the data to be produced from the extracted feature values and reduce the amount of computation needed. It identifies the strength of variations along different directions in the image data which involved computation of Eigen vectors and corresponding Eigen values. The Eigen vectors with largest associated Eigen values are the principal components and correspond to maximum variation in the data set [21].

The steps involved in PCA Include:

1. The coefficient of each iris template are converted into a one-dimensional column vector, these column vectors are stacked to form a matrix “X”.

2. The mean of each vector is given as:

\[ X_m = \frac{1}{N} \sum_{k=1}^{N} X_k \]  \hspace{1cm} (2.1)

Where \( N \) is the total number of images.

3. The mean is subtracted from all of the vectors to produce a set of zero mean vectors:

\[ X_z = X_i - X_m \]  \hspace{1cm} (2.2)

Where

\( X_z \) is the zero mean vectors.
\( X_i \) is each element of the column vector.
\( X_m \) is the mean of each column vector.

4. The Covariance matrix is computed using:

\[ C = [Xz^T \times Xz] \]  \hspace{1cm} (2.3)

5. The Eigen vectors and Eigen values are computed using:

\[ [C - \lambda I] \epsilon = 0 \]  \hspace{1cm} (2.4)

Where

\( \lambda \)’s are the Eigen value and
\( \epsilon \)’s are the Eigen vectors.

This gives \( N \) Eigen vectors \((e_1, e_2, ..., e_N)\).

6. Each of an Eigen vectors is multiplied with zero mean vector \(X_z\) to form the feature vector. The feature vector is given as:

\[ f_i = [X_z]v_i \]  \hspace{1cm} (2.5)

7. The signature of each image is found by multiplying the transpose of zero mean vectors with feature vectors given as:

\[ s_i = [Xz^T]f_i \]  \hspace{1cm} (2.6)

2.3.2 Template generation

The template that is generated in the feature encoding process, called feature codes, that is, a range of values generated from the respective iris templates by the PCA algorithm used for the iris feature extraction, was used in the identification of the different ethnic origin. These feature codes gave certain range of values when comparing templates generated from the same or similar eyes, this is known as intra-class comparisons, and another range of values when comparing templates created from different irises, known as inter-class comparisons.

3. RESULTS AND DISCUSSION

This system development focused on the extracted features of 336 iris images by the technique employed. The feature codes obtained is as shown in Table 1 where 112 African, 30 Asia1 and 19 Asia2 images were found within range 4,000,000 and 10,000,000, 82 Asia1 and 73 Asia2 iris images were within the range of
10,000,000 and 12,000,000 feature codes and 20 Asia2 images fell within range greater than 12,000,000. For more clarification, a bar chart to further explain this breakdown is shown in Fig. 1.

Table 1. Feature code ranges and the number of images generated

| Feature code (F) | Africa | Asia1 | Asia2 |
|------------------|--------|-------|-------|
| 4,000,000<F<10,000,000 | 112    | 30    | 19    |
| 10,000,000<F<12,000,000 | 0      | 82    | 73    |
| F>12,000,000      | 0      | 0     | 20    |

Fig. 1. Bar chart showing the features codes

These results showed that the dimensionality of the preprocessed images was reduced by PCA and their principal components and uniqueness were represented in form of feature codes as seen in Table 1 and Fig. 1 respectively. Even, a confirmation is seen from the detailed observation of the codes that the left and right irises of individuals are not the same. It also established that most of the people from Asia1 and Asia2 have similar texture features because they are from the same continent, Asia as their irises have medium pigmentation (located in the same 10,000,000 to 12,000,000 code region) while African irises relatively differ from the other two due to their heavy pigmentation. From the bar chart, it is noticed that African iris images were within a very low feature codes whereas highest number of iris images from Asia1 and 2 were within the same range.

A wide gap was also noticed between the region (range) where African iris images were located compared to those of Asia1 and Asia2. This showed that there exist a high difference between the African iris images and those from Asia. Also, there is a reflection of very close similarity of the Asia1 and 2 irises, that is, having been located in the same feature code subrange.

The few images that appeared for Asia alongside Africa at the first subrange explained the possibility of mixed race of subjects through inter-marriage.

4. CONCLUSION

This result showed PCA as an efficient technique for iris feature extraction as seen from the system based on the feature codes. Although this research was not performed on a very large scale of data, it has established the strength of the technique employed. Future work can explore images from the same or other ethnicities and other feature extraction techniques.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Daugman J. High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1993;1148–1161.
2. Daugman J. The importance of being random: Statistical principles of iris recognition. Pattern Recognition. 2003;36(2):279–291.
3. Roizenblatt Roberto, Paulo Schor, Fabio Dante, Jaime Roizenblatt, Rubens Belfort Jr. Iris recognition as a biometric method after cataract surgery Biomedical Engineering OnLine. 2004; 20043:2. DOI: 10.1186/1475-925X-3-2
4. Lagree S, Bowyer KW. Ethnicity prediction based on iris texture feature. Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556 USA; 2011.
5. Abbasi AA, Khan MNA, Khan SA. A critical survey of iris based recognition systems. Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan. Middle-East Journal of Scientific Research. 2013;15(5):663-668.
6. Falohun AS, Omidiora EO, Fakoluyo OA, Afolabi OA, Oke AO, Ajala FA. Development of a biometrically-controlled door system (using iris), with power backup. American Journal of Scientific and Industrial Research. 2012;3(4):203-207.
7. Falohun AS, Omidiora EO, Akanni AO, Awofadeju AS. Remote access control system using iris signature. Proceedings of the World Congress on Engineering and Computer Science. WCECS, San Francisco, USA. 2013;1:522-526.
8. Fenton S. Beyond ethnicity: The global comparative analysis of ethnic conflict. SAGE Publications. (London, Thousand Oaks and New Delhi). 2004;45(3–4):179–194. Available:www.sagepublications
DOI: 10.1177/0020715204049592
9. 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., Cash T, et al. Elsevier Press; 2012.
10. 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.
11. Qiu XC, Sun ZA, Tan TN. Learning appearance primitives of iris images for ethnic classification. International Conference on Image Processing. 2007a;2:405–408.
12. Stark L, Bowyer KW, Siena S. Human perceptual categorization of iris texture patterns. Biometrics: Theory, Applications and System, Washington, DC: BTAS; 2010.
13. Lagree S, Bowyer KW. Ethnicity prediction based on iris texture feature. Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556 USA; 2011.
14. Badejo JA, Majekodunmi TO, Atayero AA. Development of CU iris: A dark-skinned African iris dataset for enhancement of image analysis and robust personal recognition. Proceedings of the World Congress on Engineering and Computer Science, WCECS, San Francisco, USA. 2012;1.
15. Adegoke BO, Omidiora EO, Ojo JA, Falohun SA. Iris feature extraction: A survey. Computer Engineering and Intelligent Systems. 2013;4(9). ISSN 2222-1719
16. CUIRIS: Covenant University Iris (CUIRIS) iris database; 2012. Available:http://drive.google.com/CUIRISdatabase.zip
17. CASIA: Chinese Academy of Sciences Institute of Automation (CASIA) iris database; 2005. Available:http://www.cbsr.ia.ac.cn/english/irisDatabase.asp
18. CUHK: Computer Vision Laboratory, Department of Automation & Computer-Aided Engr. The Chinese University of Hong Kong, Hong Kong; 2003. Available:cnchun@acae.cuhk.edu.hk
19. Suwardi A, Takenori K, Shuhei K. Principal component analysis and sel-organizing map for visualizing and classifying fire risks in foreign regions. Japanese Society of Agricultural Informatics, Agricultural Information Research, Osaka Prefecture University, Japan. 2007;16:44-51.
20. Omidiora EO. A prototype of knowledge-based system for black face recognition system using principal component analysis and fisher discriminant algorithms, Unpublished Ph. D. (Computer Science) Thesis, Ladoke Akintola University of Technology, Ogbomoso, Nigeria; 2006.
21. Boualleg AH, Bencheriet CH, Tbeitik H. Automatic face recognition using neural network – PCA. Proceedings of the second Information and Communication Technologies Conference. 2006;1920–1925.