Asian Journal of Research in Computer Science

3(1): 1-9, 2019; Article no.AJRCOS.47947
ISSN: 2581-8260

Performance Evaluation of Feature Extraction Techniques in Multi-Layer Based Fingerprint Ethnicity Recognition System

H. O. Aworinde[1*], A. O. Afolabi[2], A. S. Falohun[2] and O. T. Adedeji[2]

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

Authors' contributions

Author HOA designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors AOA and ASF supervised and managed the analyses of the study. Author OTA managed the literature searches and review. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2019/v3i130084
Editor(s):
(1) Dr. M. Ilayaraja, Assistant Professor, Department of Computer Science and Information Technology, Kalasalingam University, Krishnankoil, Tamil Nadu, India.
Reviewers:
(1) Anthony Spiteri Staines, University of Malta, Malta.
(2) Zlatin Zlatev, Trakia University, Bulgaria.
Complete Peer review History: http://www.sdiarticle3.com/review-history/47947

Original Research Article

Received 19 December 2018
Accepted 06 March 2019
Published 19 March 2019

ABSTRACT

This paper is set out to evaluate the performance of feature extraction techniques that can determine ethnicity of an individual using fingerprint biometric technique and deep learning approach. Hence, fingerprint images of one thousand and fifty-four (1054) persons of three different ethnic groups (Yoruba, Igbo and Middle-Belt) in Nigeria were captured. Kernel Principal Component Analysis (K-PCA) and Kernel Linear Discriminant Analysis (KLDA) were used independently for feature extraction while Convolutional Neural Network (CNN) was used for supervised learning of the features and classification.
The results showed that out of sixty (60) individual fingerprints tested, eight (8) were classified as Yoruba, forty-eight (48) as Igbo and four (4) as Hausa. The Recognition Accuracy for K-PCA was 93.97% and KLDA was 97.26%. For Average Recognition time, K-PCA used 9.98seconds while KLDA used 10.02seconds. The memory space utilized by K-PCA was 94.57KB while KLDA utilized

**Corresponding author: E-mail: aworinde.halleluyah@bowenuniversity.edu.ng;*

52.17KB.
T-Test paired sample statistics was carried out on the result obtained; the outcome presented reveal that KLDA outperformed the K-PCA technique in terms of Recognition Accuracy. The relationship between the average recognition time (T_R) and threshold value (th) was found to be polynomial of order four (4) with a high correlation coefficient for KPCA and polynomial of order three (3) with a high correlation coefficient for KLDA. In terms of computation time analysis, KLDA is computationally more expensive than KPCA by reason of processing speed.

Keywords: Biometrics; deep learning; KLDA; KPCA; CNN; ethnicity; feature extraction; algorithm.

1. INTRODUCTION

Recent findings have clearly shown that despite scientific advances in computer vision, pattern recognition and machine learning, there has been various spoofing attack techniques which has been created to defeat such biometric systems [1]. Traditional Biometric systems that are based on single biometric usually suffer from problems like unacceptable error rates, spoof attacks and insufficient population coverage, expression changes, changing lightning, makeup or eye glasses and so on [2,3]. Quite a number of biometric traits can now be easily forged with the use of common apparatus and consumer electronics to imitate real biometric readings [1].

Ethnicity has however, over time been discovered not to have standard scientific definition which thereby makes this variable difficult to measure. Without a standard scientific definition, the tendency of misclassifying is very high [4]; hence, the rationale behind this research.

Deep learning is a new area of machine learning research which is about learning multiple levels of representation and abstraction that helps to make sense of data. It is a technique where many layers of information processing stages in hierarchical architectures are exploited for pattern classification and representation learning [5,6]. This work therefore, aim at evaluating the performance of feature extraction techniques for ethnicity recognition system from Deep Learning perspective.

1.2 Related Works

K. Chandra, defined ethnic identities as subset of identity categories in which eligibility for membership is determined by attributes associated with or believed to be associated with descent. Ethnicity is classified as a soft biometric trait which is meant to complement the identity information provided by the primary biometric traits such as fingerprint, face, hand geometry etc [7]. It has however been observed on many occasions and especially in this part of the world that identification of individual's ethnicity does not go beyond physical factors which in most cases can be manipulated or spoofed to carry out nefarious activities. Infact, [8] in an experiment observed that Whites of various European ethnic background usually have weak ethnic attachment especially children born to interracially married couples. From previous works on ethnicity classification [9,10,11,12], it is observed that most work done were approached using facial and Iris biometrics from the perspective of observation and machine learning; with this approach, surface level features are only captured for classification which results in high False Acceptance and False Rejection rate.

The need to authenticate individuals at various occasions cannot be overemphasized as our society is becoming more sophisticated and automated; instances of such sophistication and automation include information confidentiality, homeland security and computer security [13,14].

It has been observed that human faces provide demographic information such as gender and ethnicity. Different modalities of human faces such as range and intensity, provide different cues for gender and ethnicity identifications. Qian [8] exploited the range information of human faces for ethnicity and gender identification using a support vector machine (SVM) in which an integration scheme was also proposed for ethnicity and gender identifications by combining the registered range and intensity images. The experiments conducted on a dataset containing 1240 facial scans of 376 subjects demonstrated that the range modality provides competitive discriminative power on ethnicity and gender identifications to the intesity modality. Qian [8] in a research used 1990 Census data to examine how African-American white, Latino-White, Asian-American white, and American-Indian white couples identify their children's race/ethnicity. The study shows that choices of racial and ethnic identification of multiracial children are not as optional as for whites of various European ethnic backgrounds. They are influenced by race/ethnicity of the minority parent, intermarried couples characteristics and neighbourhood compositions. Tariq et al. [11] demonstrated gender and ethnicity identification from silhouetted face profiles using a computer vision technique where 441 images were tested. The result showed that silhouetted face profiles have a lot of information, in particular, for ethnicity identification. Shape context based matching was employed for classification. The test samples were multi-ethnic. Average accuracy for gender was 71.20% and for ethnicity 71.66%. However, the accuracy was significantly higher for some classes such as 83.41% for females (in case of gender identification) and 80.37% for East and South-East Asians (in case of ethnicity identification). Lagree and Bowyer [9] in a work examined the possibility of predicting ethnicity based on iris texture. It was stated in the work that it is possible to predict ethniticity if there are similarities in the iris texture of a certain ethnicity and these similarities differ from ethnicity to ethnicity.The authors of this work asserted that this form of soft biometric prediction could be used to narrow the search of an enrollment database for a match to probe sample. In the work, an iris image dataset representing 120 persons and 10-fold person-disjoint cross validation, 91% correct Asian/Caucassian ethnicity classification was obtained. In a related manner, a prototype video tracking and person categorization systems that uses face and person's soft biometric features to tag people while tracking them in multiple camera view was presented by Demirkus et al. [14]. The approach employed took advantage of temporal aspect of video by extracting and accumulating feasible soft biometric features for each person in every frame to build a dynamic soft biometric feature list for each tracked person in surveillance videos. Algorithms for extracting face soft biometric features were developed to achieve gender and ethnicity classification and session soft biometric features to aid in camera hand-off in surveillance videos with low resolution and uncontrolled illumination. Over 1500 face images from both genders and three ethnicity groups with various biometric algorithms were collected to train and test the face soft biometry algorithms. The algorithms achieved promising results for gender and ethnicity classification and was able to track person's re-identification for camera hand-off on low to good quality surveillance and broadcast videos.

Reviewing related works done on ethnicity, it is observed that most works done on ethnicity identification were approached from facial/iris biometrics point of view for the purpose of ethnic classification [15]; however, it is abundantly evident that in all the works reveiwed, none has vividly worked on fingerprint-based ethnicity classification which this research work intend to focus on from the perspective of Deep Learning. This paper therefore, focused on evaluating feature extraction techniques in fingerprint-based ethnicity recognition system.

2. METHODOLOGY

This section captured the overall design approach as well as the design of its internal components. The framework is a multi-level based training and classification system due to Deep Learning approach being used. As shown in Fig. 1, the model broadly encompasses the training phase and the testing phrase. The training phase comprised of fingerprint acquisition, pre-processing, feature extraction, generated template and storage while the testing phase took care of all stated in the training phase in addition with Template comparison and classification.

For extraction of features in the captured prints, Kernel Principal Component Analysis (K-PCA) and Kernel Linear Discriminant Analysis (K-LDA) were employed concurrently. The rationale behind the choice of these algorithms is the advantage they have over several other algorithms as it relates to performance optimization in high-dimensional spaces. The algorithms as well, has the ability to effectively use non-linear mappings of features (Savvides, Heo, and Park, n.d.). K-LDA as well allows efficient computation in feature space [16].

The idea of K-PCA is to first map the original input vectors x_t into a high dimensional feature space $\emptyset(x_t)$ and thereafter calculate the linear PCA in $\emptyset(x_t)$. The linear PCA in $\emptyset(x_t)$corresponds to a non-linear PCA in x_t. The PCA for x_t is calculated by

$$s_t(i) = u_i^T \emptyset(x_t) = \sum_{j=1}^{l} \alpha_i(j) K(x_j, x_t),\ i = 1, \dots, l \quad (1)$$

For making the sample input vectors in $\emptyset(x_t)$ centered $\sum_{t=1}^{l} \emptyset(x_t) = 0$ in Equation (3.1), the kernel matrix on the training set K and on the testing set K_t are modified by

$$\widehat{K} = \left(I - \frac{1}{l} 1_l 1_l^T\right) K \left(I - \frac{1}{l} 1_l 1_l^T\right), \quad (2)$$

$$\widehat{K} = (K_t - \frac{1}{l} 1_l, 1_l^T K)(I - \frac{1}{l} 1_l 1_l^T) \quad (3)$$

Where I is l-dimensional identity matrix; l_t is the number of testing data points. 1_l represents the vectors whose elements are all ones, with length l and l_t respectively. K_t represents the l_t X l kernel matrix for the testing data points.

At the training, testing and classification of generated template sections, two deep learning algorithms were adopted for that purpose. Convolutional Neural Network (CNN) serves the purpose of supervised learning of the system. The rationale behind the choice of this algorithm is due to the hybrid nature of the work being carried out in that some fingerprints will be labeled while some will be unlabeled. Equally, the algorithms are special types of multilayer neural networks and they are trained with the back-propagation algorithm.

3. RESULTS AND DISCUSSION

Both Kernel Principal Component Analysis (K-PCA) and Kernel Linear Discriminant Analysis (KLDA) which is equally known as Fishers Discriminant Analysis were used respectively for training of 1054 subjects of three major ethnic groups in Nigeria. The system was subjected to testing using Deep Learning Convolutional Neural Network algorithm from supervised learning perspective.

The approach was implemented using images normalized to 100 by 100 pixel resolutions with application of five (5) varying thresholds (0.2, 0.4, 0.6, 0.8 and 1.0) respectively in order to allow comparison between fingerprint regions. The dataset used were 1054 left and right fingerprint images of Yoruba, Igbo and Middle-belters in Nigeria.

The result gave a description of the efficiency rate of the two algorithms used for training the acquired fingerprint images. The performance of the system was evaluated using metrics such as False Acceptance Rate, False Rejection Rate, Genuine Acceptance Rate, Accuracy and Recognition Time. The results obtained from the simulation were subjected to T-Test analysis for validation purpose.

The results are shown in Tables 1 and 2 for KPCA and KLDA; the corresponding graphical representations are as well presented in Figs. 2 and 3 respectively.

The results presented in Tables 1 and 2 implied that out of sixty (60) individual fingerprints tested, eight (8) were classified as Yoruba, forty-eight (48) as Igbo and four (4) as Hausa. The Recognition Accuracy for K-PCA was 93.97% and KLDA was 97.26%. For Average Recognition time, K-PCA used 9.98seconds while KLDA used 10.02seconds. The memory space utilized by K-PCA was 94.57KB while KLDA utilized 52.17KB.

Table 1. Testing results using K-PCA

Threshold	FAR	FRR	GAR	Accuracy (%)	Recognition Time (sec)
0.2	20	3	97	91.33	9.45
0.4	16	4	96	92.00	10.12
0.6	14	4	96	92.67	9.34
0.8	10	5	95	93.33	9.89
1	4	7	93	94.00	10.02

Table 2. Testing results using K-LDA

Threshold	FAR	FRR	GAR	Accuracy (%)	Recognition Time (sec)
0.2	14	1	99	94.67	10.45
0.4	12	1	99	95.33	10.67
0.6	8	2	98	96.00	10.01
0.8	4	3	97	96.67	9.99
1	0	4	96	97.33	10.11

Table 3. Paired samples statistics

		Mean	N	Std. Deviation	Std. Error Mean
Accuracy	KLDA	96.0000	5	1.05304	.47094
	KPCA	92.6660	5	1.05462	.47164
GRR	KLDA	97.8000	5	1.30384	.58310
	KPCA	95.4000	5	1.51658	.67823
FRR	KLDA	2.2000	5	1.30384	.58310
	KPCA	4.6000	5	1.51658	.67823
FAR	KLDA	7.6000	5	5.72713	2.56125
	KPCA	12.8000	5	6.09918	2.72764

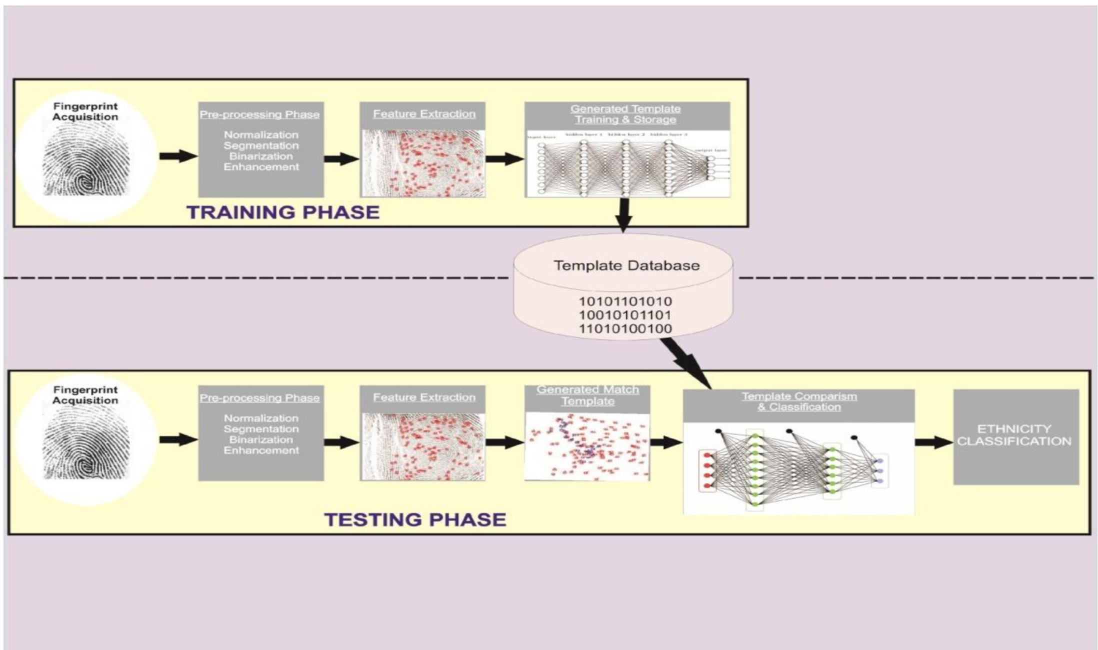

Fig. 1. Improved framework

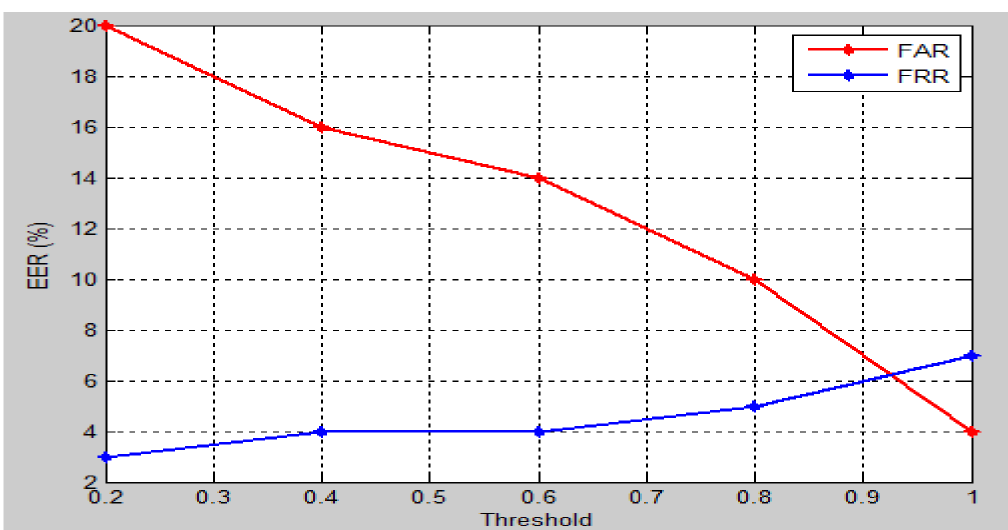

Fig. 2. Graph showing equal error rate for K-PCA

Table 4. Paired samples correlations

		N	Correlation	Sig.
Accuracy	KLDA and KPCA	5	1.000	.000
GRR	KLDA and KPCA	5	.936	.019
FRR	KLDA and KPCA	5	.936	.019
FAR	KLDA and KPCA	5	.985	.002

4. PERFORMANCE EVALUATION

The results obtained from the simulated model were used to evaluate the performance of the system in order to validate its outcome and as well compare the two algorithms used. T-Test paired sample statistics was carried out on the result with focus on Accuracy of the system, Genuine Rejection Rate and False Rejection Rate. The outcomes are presented in Tables 3, 4 and 5 respectively.

H_0: The difference between the accuracy of KLDA and KPCA is not statistically significant.
H_1: The difference between the accuracy of KLDA and KPCA is statistically significant.

The paired t-test analysis conducted between the accuracies of KLDA and KPCA at different threshold value reveal that there is no much distinction in the test result with mean difference of 3.33 (i.e. $\mu = 3.33$). Nevertheless, the result confirmed that the KLDA is statistically significant at $P < 0.01; P = 0.000$ with $t\ value = 1361.1$. The t-test result validates the fact the KLDA outperformed the KPCA techniques in terms of recognition accuracy. Therefore, the alternative hypothesis which states that the difference between the accuracy of KLDA and KPCA is statistically significant is accepted.

H_0: The difference between the GRR of KLDA and KPCA is not statistically significant.
H_1: The difference between the GRR of KLDA and KPCA is statistically significant.

The paired t-test analysis conducted between the GRR of KLDA and KPCA at different threshold value shows that there is slight difference in the test result with mean difference of 2.4 (i.e. $\mu = 2.4$). Nevertheless, the result confirmed that the KLDA is statistically significant at $P < 0.01; P = 0.001$ with $t\ value = 9.798$. The t-test result validates the fact the KLDA outperformed the KPCA techniques in terms of GRR. Therefore, the alternative hypothesis which states that the difference between the GRR of KLDA and KPCA is statistically significant is accepted.

H_0: The difference between the FRR of KLDA and KPCA is not statistically significant.
H_1: The difference between the FRR of KLDA and KPCA is statistically significant.

A t-test value was measured between the FRR of KLDA and KPCA. The paired t-test analysis conducted reveals that KLDA was statistically significant at $P < 0.01; P = 0.001$ with $Mean\ difference = -2.4, df = 4\ and\ t\ value = -9.798$. The mean difference and t-value being negative assert the fact the KLDA have a reduced False Rejection Rate. The t-test result validates the fact the KLDA outperformed the KPCA techniques in terms of FRR. Therefore, the alternative hypothesis which states that the difference between the FRR of KLDA and KPCA is statistically significant is accepted.

H_0: The difference between the FAR of KLDA and KPCA is not statistically significant.
H_1: The difference between the FAR of KLDA and KPCA is statistically significant.

A t-test value was measured between the FAR of KLDA and KPCA. The paired t-test analysis conducted reveals that KLDA was statistically significant at $P < 0.01; P = 0.000$ with $Mean\ difference = -5.2, df = 4\ and\ t\ value = -10.614$. The mean difference and t-value being negative assert the fact the KLDA have a reduced False Acceptance Rate. The t-test result validates the fact the KLDA outperformed the KPCA techniques in terms of FAR. Therefore, the alternative hypothesis which states that the difference between the FAR of KLDA and KPCA is statistically significant is accepted.

$$T_R = 48.018th^3 - 71.793th^2 + 26.791th + 0.8217 \qquad R^2 = 0.8599 \tag{4.1}$$

$$T_R = -3.7675th^2 + 5.99 + 1.5441 \qquad R^2 = 0.9232 \tag{4.2}$$

Table 5. Paired samples test

		Paired Differences					T	Df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Accuracy	KLDA - KPCA	3.33400	.00548	.00245	3.32720	3.34080	1361.100	4	.000
GRR	KLDA - KPCA	2.40000	.54772	.24495	1.71991	3.08009	9.798	4	.001
FRR	KLDA - KPCA	-2.4000	.54772	.24495	-3.08009	-1.71991	-9.798	4	.001
FAR	KLDA - KPCA	-5.2000	1.09545	.48990	-6.56017	-3.83983	-10.614	4	.000

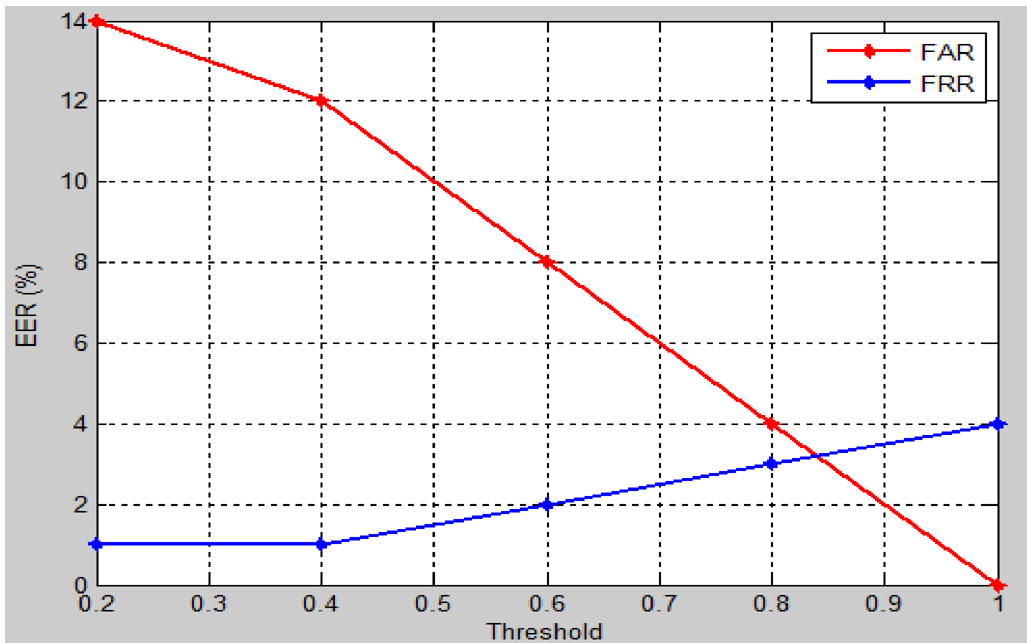

Fig. 3. Graph showing equal error rate (EER) for K-LDA

In view of the above computation time analysis, KLDA is more computationally expensive than KPCA in terms of processing speed and memory usage in ethnicity fingerprint recognition system.

5. CONCLUSION

In this work, the performance of two different feature extraction algorithms (K-PCA and K-LDA) were evaluated in order to ensure the efficiency of the system. It was however discovered that K-LDA outperforms K-PCA in terms of Average recognition time and recognition accuracy; however, K-LDA is computationally more expensive than K-PCA. This work will in no small measure contribute to existing knowledge in pattern recognition through the fusion of soft biometric trait (ethnicity) with hard biometric feature (fingerprint) for the purpose of enhancing the process of identification through machine. Since the performance of the two techniques (K-LDA and K-PCA) have been substantiated in this work, the multiclass variant of K-LDA can be modified to enhance the process of feature extraction and thereafter subject such features to unsupervised deep learning.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Menotti D, Chiachia G, Pinto A, Schwartz WR, Pedrini H, Falc AX, Rocha A. Deep representations for iris, face, and fingerprint spoofing detection. IEEE Transactions on Information Forensics and Security. 2015;10(4):864–879.
   DOI:http://doi.org/10.1109/TIFS.2015.2398817
2. Afolabi A, Ademiluyi D. Performance evaluation of some selected feature extraction algorithms in ear biometrics. International Journal of Science and Research. 2015;4(3):2384–2390.
3. Zewail R, Elsafi A, Saeb M, Hamdy N. Soft and hard biometrics fusion for improved identity verification. In The 2004 47th Midwest Symposium on Circuits and Systems, 2004. MWSCAS '04. (Vol. 1, p. 1_225-1_228). IEEE.
   DOI:http://doi.org/10.1109/MWSCAS.2004.1353967
4. Kaneshiro B, Geling O, Gellert K, Millar L. The challenges of collecting data on race and ethnicity in a diverse, multiethnic state. Hawaii Medical Journal. 2011;70(8):168–71.
   Available:http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3158379andtool=pmcentrezandrendertype=abstract
5. Deng I. A tutorial survey of architectures, algorithms, and applications of Deep Learning. APSIPA Transactions on Signal and Information Processing; 2014.
   DOI:https://doi.org/10/1017/astip.2013.9
6. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444.
   DOI:https://doi.org/10.1038/nature14539
7. Jain AK, Dass SC, Nandakumar K. Soft biometric traits for personal recognition systems. Proc. of International Conference on Biometric Authentication (ICBA), (July), 2004;731–738.
   DOI:http://doi.org/10.1007/978-3-540-25948-0_99
8. Qian Z. Options: Racial/Ethnic Identification of children of intermarried couples. Social Science Quarterly, 2004;85(3):746–766.
   Available:http://isites.harvard.edu/fs/docs/icb.topic548257.files/Options Racial Ethnic ID of Children of Intermarried Couples.pdf
9. Lagree S, Bowyer KW. Ethnicity Prediction Based on Iris Texture Features; 2007.
10. Lu X, Chen H, Jain AK. Multimodal facial gender and ethnicity identification. Advances in Biometrics; 2005.
11. Tariq U, Hu Y, Huang TS. Gender and ethnicity identification from silhouetted face profiles. In Proceedings - International Conference on Image Processing, ICIP; 2009.
    DOI:http://doi.org/10.1109/ICIP.2009.5414117
12. Zhang D, Lu G. 3D Biometrics; 2013.
13. Orike S, Anireh VIE, Ibironke AS. A gender and ethnicity identification system in Nigeria using the Fingerprint Technology. Proceedings of the World Congress of Engineering, I; 2016.
14. Demirkus M, Garg K, Guler S. Automated person categorization for video surveillance using soft biometrics. Proceedings of the SPIE. 2010;7667. DOI:http://doi.org/10.1117/12.851424

15. Latinwo BM, Falohun AS, Omidiora EO. Principal component analysis-based ethnicity prediction using iris feature. British Journal of Applied Science and Technology. 2016;16(4):1-5.
16. Mikat S, Fitscht G, Weston J. Scholkopft, Bernhard, and Mullert, K.-R. (n.d.). Fisher Discriminant Analysis with kernels. Available:http://courses.cs.tamu.edu/rgutier/csce666_f16/mika1999kernelLDA.pdf

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle3.com/review-history/47947