Similarities in African Ethnic Faces from the Biometric Recognition Viewpoint

Ogechukwu Iloanusi
University of Nigeria
Department of Electronic Engineering
ogechukwu.iloanusi@unn.edu.ng

Patrick J. Flynn
University of Notre Dame
Department of Computer Science and Engineering
flynn@nd.edu

Patrick Tinsley
University of Notre Dame
Department of Computer Science and Engineering
ptinsley@nd.edu

Abstract

Face pose, illumination, and facial expressions are known factors that affect face recognition performance and have been studied at length in the literature. The impacts of demographic factors such as gender, race, and age on performance have also been studied, with increasing interest recently in the context of algorithmic bias concerns. This work is a study of face recognition performance using a database of faces of Nigerian subjects drawn from 28 different ethnicities. There are documented differences in facial anthropometric characteristics between some Nigerian ethnicities, and this study was intended to establish initial results regarding the impact of these inter-ethnic differences on face recognition performance. A comparison to performance on a database of Caucasian/Asian face images is made. Our study analyzes how 28 African ethnicities affect face identification performance metrics by focusing on the genuine and impostor scores' distributions. Our analysis shows that face identification performance is not remarkably influenced by varying ethnicities within the African race though there are significant differences in relation to the Caucasian/Asian set.

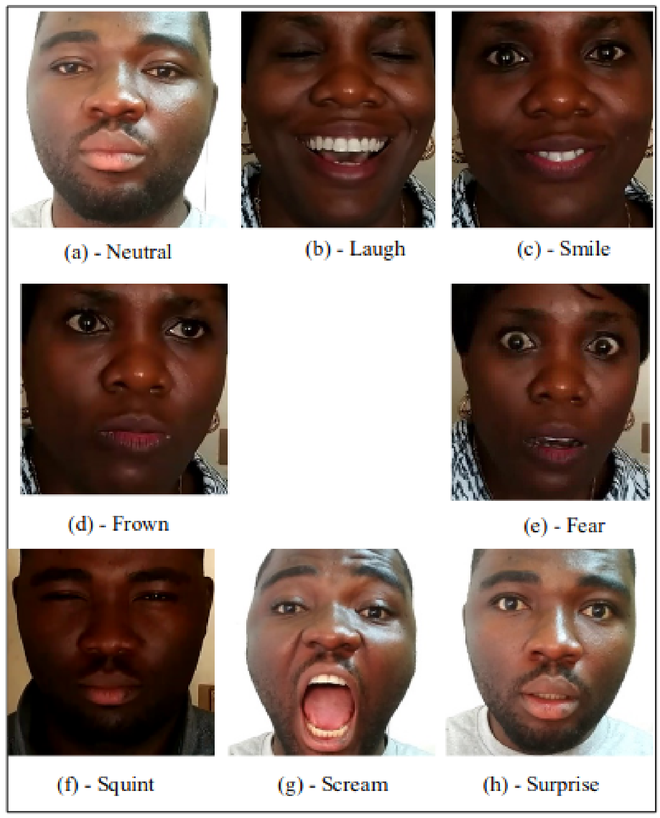

Figure 1. Examples of eight facial expressions in the *Nigerian Faces* data set.

1. Introduction

Race and ethnicity are closely related but distinct concepts. Race is a biological or genetic construct. Ethnic groups are often distinguished by race, but also by language, geography, customs, and culture. Race and ethnicity have varying meanings across Central America, Latin America, Africa and Asia.

Nigeria is a multi-lingual, multicultural country with over 250 ethnicities (or tribes) across its 36 states. Members of specific ethnicities generally identify very strongly with that ethnicity. Exogamous marriages in Nigeria are quite uncommon [23, 33], which fosters conservation of genetic factors and cultural practices. The ethnic groups in Nigeria are diversified and can be described by their culture and language, but also by head and facial morphology to an extent. The head/facial attributes can manifest in extrinsic features such as tribal marks [2, 39], but also in intrinsic features such as facial symmetry or proportions, which can be quantified in facial anthropometric measurements [4, 17, 32, 37, 41]. The facial features that make this classification easy for humans are not that obvious and at times difficult to define. However, there are many studies emphasizing that there are genetic factors influencing face and body morphology amongst races globally [8, 12, 20, 26, 30, 40, 43, 44, 45, 47]. A deep learning-based classifier was trained in [22] to classify faces as Asian, African, Indian and Caucasian; further ethnicity classification was performed in [35]. There have been several works that have studied the facial anthropometry of various tribes in Nigeria and found similarities in anthropometry within tribes and dissimilarities between the studied tribes [32, 37, 41, 17, 4].

In this paper, we analyze the performance of face recognition algorithms on gallery and probe sets composed of different faces of African and/or Caucasian ethnicities. We demonstrate this by using an acquired database of African faces (*Nigerian Faces*) and a Caucasian dataset (*ND Faces*), and by defining four major experiments that are carried out on these data sets. The contributions of this paper are as follows:

- A first study on the influence of Nigerian ethnicities on face biometric identification performance.
- A comparative study of genuine and impostor match scores between African faces (from *Nigerian Faces*) and Caucasian/Asian faces (from *ND Faces*).
- The *Nigerian Faces* data set, available to researchers, to facilitate development of and reduction of bias in face recognition systems for persons of African heritage [22]. The *ND Faces* data set will also be made available to researchers.

2. Review of Related Work

2.1. Related Work on Genetic and Anthropometric Factors Influencing Facial Morphology

The genetic and anthropometric factors influencing facial morphology have been studied to a great extent in [8, 12, 20, 26, 30, 40, 43, 44, 45, 47]. Studies on different races from the genetic point of view show that some facial measurements are heritable [7, 34]. A study of 229 healthy individuals from 38 Korean families [26] explores the dependency of facial morphology and anthropometric traits on genetic factors. A large-scale heritability study of 952 female British twins were carried out in [40] with the aim of characterizing facial geometry at a fine level. The authors in [8] devised models for quantifying facial traits for forensic, clinical and epidemiological purposes. A study of the genetic traits influencing facial and other anthropometric traits in over 3,500 Bantu African children and adults and over 2,400 Europeans in [12] show that facial traits are most heritable, particularly horizontal measurements in the face and face circumference. The genome-wide associations responsible for the replication of the facial traits were identified in Africans and found to be different from the genome-wide associations in Europeans in [12]. In [44], a bioinformatics analysis of the genetic variants influencing craniofacial morphology was conducted via whole-exome sequencing amongst 50 northern Han Chinese participants for the purposes of understanding human growth and phenotyping in humans. The genes affecting facial morphology in European population was also studied in [30]. The authors in [45] studied the diversity of faces from the genetic viewpoint. The authors in [43] reviewed the successes and challenges in studying the genes influencing facial morphology. Facial features from populations from three African countries, 169 Ugandans, 234 Nigerians and 110 Kenyans, were studied in [36] and found to be significantly different with respect to their face height and nasal length.

Within Nigeria, the study of the Bini tribe in [32] addressed a certain nasofacial anthropometric tendency (noses tend to be platyrrhine and faces tend to be mesoproscopic). Photometric facial analysis of the adult Igbo male conducted in [41] showed that the males had a less prominent chin and a projected nose. The facial anthropometry of Urhobos was studied in [17], and the results showed that Urhobo faces are generally mesoproscopic and the nasal height in males is higher than that of females. The study of the facial anthropometry of some adult Tiv and Idoma tribes in [37] showed that the Idoma faces studied are leptoproscopic, while the Tiv faces are mesoproscopic. The study of 560 subjects of the major ethnic groups –Igbo, Hausa and Yoruba in [4] showed that there were quantifiable anthropometric parameters that differentiate these ethnic groups. Cephalometric indices were calculated from the anthropometric measurements and studied using Analysis of Variance (ANOVA). The clustering of the features in the clustering space showed a clear separation of clusters for the ethnic groups Igbo, Hausa and Yoruba in [4] – confirming that the features are discriminative and the differences were attributed to genetic factors. A study of the Igbo, Hausa and Yoruba tribes in [3] also shows that the faces are anthropometrically different. Genetic variations amongst the Igbo, Hausa and Yoruba ethnicities were studied in [38] and found to vary amongst the three tribes. The anthropometric nasal features for 385 Hausa and 197 Yoruba individuals were studied in [6] and found to be significantly different between the two ethnicities. The face and nose shape for the Ukwuani tribe was studied in [18] and found to be hypereuryprosopic and platyrrhine, respectively. The study in [25] shows that the nasal indices in males are higher than in females. Facial biometrics of Yorubas were studied in [5] and found to be different for the males and females. These many works confirm that facial morphology varies amongst African ethnicities.

2.2. Related Work on Studies of Various Ethnic Groups or Races on Biometric Recognition Performance

There have been very few works in the literature that studied the effect of race and/or ethnicity on genuine and/or impostor match score distributions. The work in [29] characterised the effects of race on face verification accuracy by analysing the false match (FMR) and false non-match rates (FNMR), genuine and impostor distributions in African-American and Caucasian face images in the MORPH database. The differences in the genuine and impostor distributions were found to vary significantly. The work also reveals a high FMR with African-American face verification. In [21], the effects of age, race, and gender were studied by analysing the FNMR and impostor distributions based on information entropy. Race was identified to influence the FNMR more than other two factors from its higher information gain.The work in [28] studies the separate influence of race and skin tone on biometric identification performance of African-American and Caucasian face images. It adds to the evidence that that face recognition accuracy varies for African-Americans and Caucasians and their results show an increase in false match rate for African-Americans and false non-match rate for Caucasians.

The work in [1] studies the effects of population demographics (namely, race, age and gender) on the performance of face recognition systems. Majority race training effects were identified which led to better training performance when one race is used and degraded performance when multi-racial groups are used for training, as well as significantly varying performances in a given algorithm on evaluation on databases of different races. In [9], a detailed study on the MORPH dataset reveals why the accuracy of face recognition algorithms depend on race despite the novelties of recent deep face models. Strategies that could be valuable in improving face recognition and limiting "other race effects" were studied in [11]. This could be beneficial in developing robust and unbiased face recognition algorithms. In [10], the effects of several factors (demographics, operational threshold decisions, image quality, and algorithm) on face recognition accuracy were studied on East Asian and Caucasian face data.

Table 1. Distribution of Subjects in the 28 Ethnicities (Afemai through Yoruba) in the *Nigerian Faces* Data Set

Groups	Subject Ethnicity	Samples
Mixed Tribes	Afemai (2), Akoko-Edo (1), Anang (3), Efik (2), Ekoi (1), Fulani (2: Expressions–1 / Poses-2), Hausa (2), Ibibio (7), Idoma (4), Igala (7), Igede (1), Ijaw(2), Ika (1), Ikom (1), Ikwerre (2), Ishan (2), Kanuri (4), Karaikarai (1), Obudu (1), Ogoni (1), Okirika (1), Okobo (1), Tiv (3), Ukwani (2), Urhobo (5), Yako (1), Yoruba (2)	4,035
One Single Tribe	Igbo (489: Expressions – 471 / Poses – 489)	30,915
Sum Total	551	34,950

3. Data

3.1. Nigerian Faces

Face videos were acquired from a total of 551 subjects (356 males and 195 females) from 28 distinct ethnicities. Face videos with both expression and head pose variations were collected in two major sessions between June and November, 2018. 534 subjects participated in the first session. In the second session, 151 subjects participated; 134 subjects participated in both sessions. Since the dataset was collected in two sessions, subjects naturally appeared different across the sessions, resulting in changes of clothing, makeup, hair style or haircut, emotions, etc. Seventeen more subjects of one ethnicity (9 males and 8 females) participated in a series of collection dates between April and December 2019. Sixteen of them had only their poses acquired and one had her facial expression acquired. All face images in this data set were collected from consenting subjects under a human subjects research protocol approved by the Ethics Committee of University of Nigeria.

The data was collected in indoor and outdoor environments using a mobile phone camera. The outdoor environment was only used for data collection in the first session. Photos have an aspect ratio of 16:9. Video recording was set to 1920×1080 at 30fps. Most of the videos were collected under challenging conditions, (e.g., interference with objects in the scene, acquisition in the evening, dark backgrounds, imaging under a canopy, in open air and under too bright sunlight).

The high-resolution face dataset is comprised of two curated face videos: face expressions and face poses, acquired for each of the subjects. The face videos were captured at distances within 0.5 meter to 1.2 meters. The facial expressions include the following conditions: neutral, laughing, smiling, frowning, fearful, squinting, screaming, and surprised. Cropped images of eight facial expressions are shown in Figure 1 (a) through (h). The facial poses include frontal, elevated or lowered pitch, and right or left yaw. Subjects were given instructions on the expressions and poses to portray before filming.

Persons from twenty-eight distinct ethnicities participated in the face data collection. Face images acquired from the 28 ethnicities were grouped into two categories in Table 1. The numbers in parenthesis under "Ethnicities" in Table 1 show the total number of subjects per ethnicity. The number of subjects acquired in the Expressions and Poses datasets are the same in the majority of ethnicities, except in the cases of the Fulani and Igbo ethnicities. The number of acquired Fulani subjects are 1 and 2 in the Expressions and Poses sets, respectively. The number of Igbo subjects are 471 and 489 in the Expressions and Poses sets, respectively. Acquiring data from over 300 ethnicities in Nigeria would require a tour of all 36 States and ethnic groups in the country as ethnic groups in Nigeria tend to be localized; the relative populations of these groups also varies greatly. As such, many of the ethnicities in Nigeria are not sampled, and most of those that were sampled yielded a small number of subjects. As a result, the data set is quite imbalanced, with one ethnicity (Igbo) accounting for nearly 90% of images acquired. This imbalance drove the design of our experiments. Still images in the data set were extracted from videos at a rate of one frame per second. In our experiments, images with non-frontal head pose were excluded. As a result, the *Nigerian Faces* data set used in the experiments below comprises 471 Igbo subjects and 62 subjects from a mixture of 27 non-Igbo ethnicities, totalling 533 subjects. Each subject has between 10 and 54 facial images in the gallery or probe set, totalling 15,394 between experiments. Additional subject data was collected in addition to ethnicity, such as parental ethnicity, weight, height, and age, but were not used during experiments.

3.2. ND Faces

A counterpart Caucasian/Asian face image data set (*ND Faces*), was similarly analyzed in this work, though less intensively. The data set was collected at the University of Notre Dame across seven different recording sessions and includes 7,728 high resolution frontal face images of 610 total subjects (313 males and 297 females). Though collected across different sessions, all images feature persons in uncontrolled, indoor environments and were captured using the same camera (Nikon D90) at 4288×2848 resolution. However, unlike the *Nigerian Faces* data, the *ND Faces* images were recorded as stills, as opposed to snapshots from video footage. All face images in the *ND Faces* data set were collected from consenting subjects under a human subjects research protocol approved by the University of Notre Dame Human Subjects Institutional Review Board.

The ND Faces data set contains 479 subjects who identified as "White" while the remaining 131 identified as "Asian". Within the "Asian" subset were the following: Asian-Unspecified (95), Asian-Southern (11), Asian-Indian (8), Asian-Chinese (7), Asian-Middle-Eastern (4), Asian-Filipino (3), Asian-Korean (2), and Asian-Vietnamese (1).

4. Experiments

4.1. Experimental Setup

The experiments were carried out in Python and MATLAB environments. The ArcFace matcher [13] was used to extract 512-element feature vectors from cropped and resized face chips extracted from the database images.

There are four face recognition experiments in this paper, and are so numbered 1 through 4. Experiment 1 assesses recognition performance within a single ethnicity (part A) and across a multi-ethnic data set (part B). Experiments 2A and 2B build on experiments 1A and 1B by adding a set of impostors to the gallery in both parts (respectively), keeping other elements of the experiments identical. Experiment 3, by contrast, compares impostor matches only, from the previously described data sets of African and Caucasian/Asian subjects. Experiment 4 compares impostor and genuine match scores for Caucasian/Asian, African, and combined Caucasian/Asian and African face images.

- Experiment 1 is conducted for 2 subsets of the *Nigerian Faces* data set:
  - A mixture of 27 tribes (all non-Igbo subjects)
  - A single tribe (Igbo)

Gallery and probe sets are drawn from subjects within each group.

- Experiment 2 is conducted for the following:
  - A mixture of 27 tribes (non-Igbo), including an expanded impostor set of images of Igbo subjects
  - A single tribe (Igbo), including an expanded impostor set from the same tribe

Probe sets in 2A and 2B are similar to those in 1A and 1B. The gallery sets include an expanded impostor set of one tribe.

The goal of the first two experiments are to compare verification performances for similar experiments that differ in the number of impostor comparisons. In experiments 2A and 2B, the number of subjects and gender in the gallery and probe sets are equal. There is consistency in the number of subjects and gender.

Experiment 3 and Experiment 4 involve studying biometric match scores from the Nigerian Faces data set in relation to the *ND Faces* data set. Experiment 3 analyses impostor pairings between the face images in each of the two databases, while experiment 4 analyzes genuine and impostor pairings completely within the two databases.

4.2. Nigerian Faces Gallery and Probe Sets

The gallery and probes sets for experiment 1A–mixed tribes, experiment 1B–Single tribe, experiment 2A-Mixed tribes, and experiment 2B–Single tribe, are defined in this section. The gallery and probe sets notations; number of subjects, samples, male and female gender in the gallery and probe sets are defined in Table 2. Consistency was maintained in the gallery and probe sets across all four experiments. Subjects-wise, the gallery and probe sets are equal for experiment 1 and experiment 2. The consistency is maintained subject-wise, and not by number of images, since inter-class variations are expected to be much higher than intra-class variations in biometrics recognition.

In Experiments 1A and 1B, there are 62 subjects in the gallery, $\mathrm{M_G}$, $\mathrm{S_G}$ and 18 subjects in the probe set, $\mathrm{M_P}$, $\mathrm{S_P}$ for both experiments. The numbers of males and females are also equal across subjects in both experiments. All probe images have matching images in the gallery across both experiments too, i.e. $\mathrm{M_G} \cap \mathrm{M_P} = \mathrm{M_{PG}}$ and $\mathrm{S_G} \cap \mathrm{S_P} = \mathrm{S_{PG}}$.

Experiment 2 includes an impostor set $\mathrm{E_G}$ of 409 subjects which are not part of Experiment 1. This impostor set is used for expanding the gallery set of Experiment 2, $\mathrm{ME_G}$, $\mathrm{SE_G}$ to a subject size of 471 in Experiments 2A and 2B. The expanded gallery set are thus: $\mathrm{ME_G} = \mathrm{E_G} \cup \mathrm{M_G}$ and $\mathrm{SE_G} = \mathrm{E_G} \cup \mathrm{S_G}$, for Experiments 2A and 2B, respectively. The probe set in experiments 2A and 2B, $\mathrm{M_P}$, $\mathrm{S_P}$ still have 18 subjects each. All probes in each of the Experiment 2 have matching impressions in their corresponding gallery set, i.e., $\mathrm{ME_G} \cap \mathrm{M_P} = \mathrm{M_{PG}}$ and $\mathrm{SE_G} \cap \mathrm{S_P} = \mathrm{S_{PG}}$.

4.3. Experimental Methods

4.3.1 Experiment 1 Methods

In Experiment 1A, face identification experiments were carried out between S_G and S_P, as defined in the (respective) gallery and probe sets for this experiment in Table 2. Face identification experiments were carried out between M_G and M_P in Experiment 1B.

4.3.2 Experiment 2 Methods

The objective of Experiment 2 is to determine if performance is undermined or not when an impostor set is present in the gallery (containing a different ethnicity for experiment 2A, and the same ethnicity for Experiment 2B).

It is a known fact that the gallery size affects the identification error rate [16, 19, 24]. The error increases with the database size. EER depends on the size of both probes and enrolled templates. Hence it is expected that the errors will increase for comparisons involving more references and probes, i.e. more match scores.

We determine if the combination with impostors of other ethnicities impacts or improves performance. The objective is to evaluate if there are ethnicity-based factors that influence the score distributions. This can be evaluated from the d' score. A significant increase in d' would mean that the faces in the genuine set are characteristically different, in a way that manifests in the feature vectors, and are hence reflected in the separation of the genuine and impostor distributions.

Table 2. Probe and Gallery Sets for All Experiments

	Subjects			**Samples**
Experiment 1	***Total***	***Female***	***Male***	
Mixed Tribes, M				
Gallery, $\mathrm{M_G}$	62	24	38	1,348
Probe, $\mathrm{M_P}$	18	12	6	536
$\mathrm{M_{PG}}$	18	12	6	1,020
Single Tribe, S				
Gallery, $\mathrm{S_G}$	64	24	38	1,470
Probe, $\mathrm{S_P}$	18	12	6	538
$\mathrm{S_{PG}}$	18	12	6	983
Experiment 2	***Total***	***Female***	***Male***	
Expanded Impostor Gallery Set, $\mathrm{E_G}$	409	194	260	8,917
Mixed Tribes, M				
Gallery, $\mathrm{ME_G}$	471	173	298	10,265
Probe, $\mathrm{M_P}$	18	12	6	536
$\mathrm{M_{PG}}$	18	12	6	1,020
Single Tribe, S				
$\mathrm{SE_G}$	471	173	298	10,387
Probe, $\mathrm{S_P}$	18	12	6	538
$\mathrm{S_{PG}}$	18	12	6	983

4.3.3 Experiment 3 Methods

Experiment 3 included an analysis of impostor score distributions between two ethnic composite groups: African and Caucasian/Asian. After extracting ArcFace feature vectors for all images across both data sets, a set of impostor image pairs (featuring images from different subjects) was constructed. Since each of the 537+619 subjects in both data sets have several images, a full impostor set including all impostor pairs would contain more than 100 million scores. To avoid computing the full impostor set, we took a 100-subject subset of each dataset, resulting in about 3.93 million impostor pairs and corresponding impostor scores; we believe this number is large enough to properly approximate the real distribution found in the full impostor set. The match score distribution results are plotted in Figure 4 in red (+ markers).

4.3.4 Experiment 4 Methods

Experiment 4 provides analyzes of both genuine and impostor score distributions within the African and Caucasian/Asian groups listed above. Using extracted ArcFace features, genuine image pairs and scores (featuring the same subject) were computed for both races. The results (about 319,000 and 79,000 scores, respectively) for African and Caucasian/Asian were plotted in Figure 4, and are labelled in green square and green diamond makers, respectively.

Impostor scores, similar to those detailed above in Section 4.3.3, were also computed in a similar fashion: using a 100-subject subset from each race and computing approximately 7.58 and approximately 1.66 million impostor match scores, respectively. These results were plotted in Figure 4 in red X markers, thick and thin respectively.

4.3.5 Face Detection Yield Methods

In addition to generating ArcFace feature-based match scores, an analysis of face detection yields was conducted for all face images in the *Nigerian Faces* and *ND Faces* data sets. This experiment was motivated by the possibility of a statistically significant difference in face detection yield between races, which has proven to be an emergent and topical theme in recent research. Four face detectors were used to ingest all expression-based images from each data set: (1) Haar cascades [42], an older CPU-based face detection schema, (2) dlib [27], a state-of-the-art CNN-based face detector, (3) MTCNN [46], a multi-task cascaded convolutional network face detector, and (4) insightface's RetinaFace [15], another CNN-based face detector. In total, there were 15,934 images for the Nigerian Faces data set, and 7,728 images for the *ND Faces* data set. The results of these face detect experiments are listed in Table 3.

4.4. Performance Metrics

All possible match scores were obtained from an exhaustive comparison of all face feature vectors between probe and gallery samples in experiments 1 and 2 by taking the L_2 norm. Face matching was based on similarity scores obtained from the reciprocal of the dissimilarity scores.

The genuine histogram, impostor histogram, genuine probability density function (pdf), impostor pdf, equal error rate (EER) and area under the ROC curve (AUC) were obtained from the sets of genuine and impostor similarity scores. A d' score was obtained as determined in [31].

5. Results

The results of experiments 1 and 2 portray some or all of the variables in the following terms:

Full Name	**Abbreviated Form**
Total genuine / impostor scores	Gscores# / Iscores#
Mean genuine / impostor score	mean_Gs/ mean_Is
Max genuine / impostor score	max_Gs/ max_Is
Minimum genuine / impostor score	min_Gs/ min_Is
Area under the ROC curve	AUC
Equal error rate(%)	EER

5.1. Experiment 1 Results

Results of the two face identification experiments in experiment 1A and 1B conducted using the gallery and probe sets in Table 2 are presented in Figure 2.

- In experiment 1A, depicted in Figure 2 (i), there are 14,391 genuine and 708,137 impostor scores. The AUC and EER are 0.9975 and 0.4028%, respectively. The minimum genuine and maximum impostor scores are 44.6823 and 64.8361, respectively. The d' score is 3.8322.
- In experiment 1B, depicted in Figure 2 (ii), there are 12,811 genuine and 778,049 impostor scores. The AUC and EER are 1.000 and 0.1738%, respectively. The minimum genuine score and maximum impostor score are 44.9001 and 75.1964, respectively. The d' score is 3.8437.

Experiment 1 forms the basis for experiment 2. From the results of experiment 1, we see that the d' score is not a function of increasing or decreasing gallery sets. Although $ME_G >> M_G$, the d' scores of 3.8322 and 3.8437, for experiments 1A and 1B (respectively) are almost equal. In both results, the average genuine scores are well above the average impostor scores. In effect, perhaps a few impostor scores could have exceeded the minimum genuine score.

5.2. Experiment 2 Results

Experiment 2 consisted of two experiments, each employing an expanded gallery set of a single tribe, as defined in Table 2, thus increasing the number of impostor scores over the number encountered in experiment 1. Results of face identification experiments conducted on these sets are shown in Figure 3. In experiment 2A in Figure 3 (i), the combination of a gallery set of 27 mixed tribes with a 28th tribe does not drastically change when compared to the gallery set of 27 tribes in experiment 1A; the

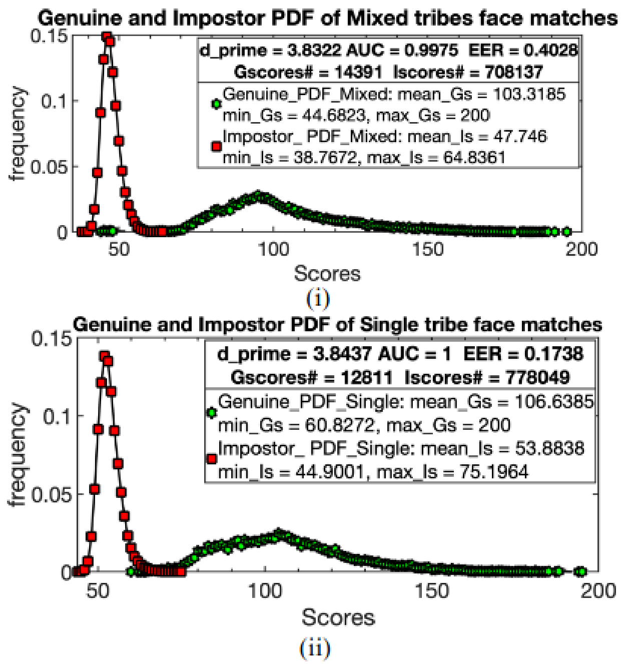

Figure 2. Comparison of probes and gallery sets in groups of Mixed Tribes (i) and Single Tribes (ii)

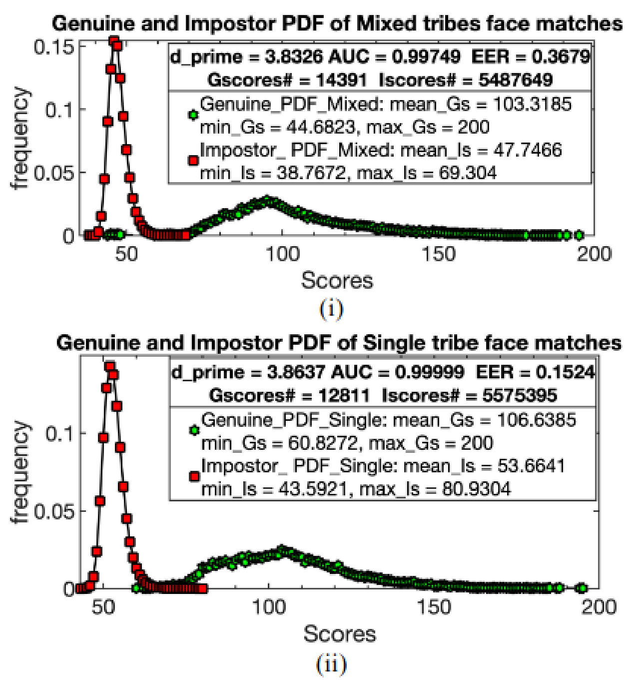

Figure 3. Comparison of probe and expanded gallery sets in groups of Mixed Tribes (i) and Single Tribes (ii)

d' scores are 3.8322 (as seen in Figure 2 (i)), and 3.8326. Since the probe sets are the same in experiment 1A and 2A, the mean and minimum genuine scores statistics are unaffected in both experiments. The mean impostor scores in

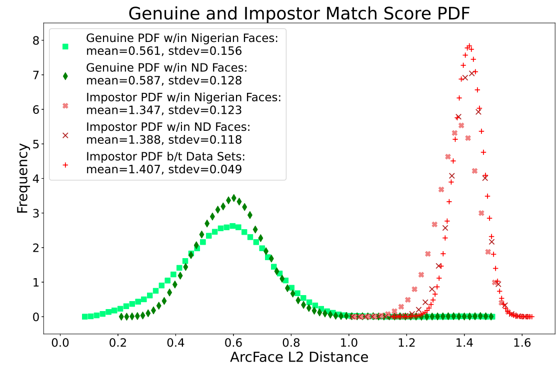

Figure 4. Genuine and impostor ArcFace match scores for the *Nigerian Faces* and *ND Faces* data sets.

both experiments are almost the same, 44.746 and 44.7466, in 1A and 2A, respectively. This shows that the introduction of an impostor set from another tribe does not affect the impostor population distribution. The results for the experiments shown in Figure 3 (ii) conducted on a single tribe using an expanded gallery of a single tribe in experiment 2B is better than that of experiment 1B with a smaller gallery set.

The statistics of the genuine scores' distribution in experiments 2A and 2B are unchanged. Additionally, despite adding a larger impostor set, the impostor score distributions remain relatively unchanged as well, as evidenced by d' values (3.8437 versus 3.8637).

5.3. Experiment 3 Results

As seen in Figure 4's pdf's (red plus markers), the cross-race impostor scores between the Nigerian Faces and *ND Faces* face images are among the largest in terms of L_2 distance. During the process of training ArcFace [14], similar faces are awarded lower distances and dissimilar faces are awarded larger distances. Consequently, this larger average distance between ArcFace features suggests that Caucasian/Asian face images and African face images tend to exhibit pronounced differences, at least in the context of biometric facial recognition. However, when combined with the results from the next section, we see that this cross-race impostor distribution is not unlike corresponding same-race distributions.

5.4. Experiment 4 Results

As discussed previously in Section 4.3.4, our fourth set of experiments dealt with genuine and impostor score distributions completely within each of the races in our respective data sets. These four match score/distance distributions were computed and are shown in Figure 4. In green markers are pdf's showing genuine (same-subject) match

Data Sets	Haar Cascades	dlib	MTCNN	Retina Face
Nigerian Faces	0.9185	**0.9976**	0.9955	0.9936
ND Faces	0.9992	0.9930	0.9720	**0.9997**

Table 3. Face detection yields (%) across both data sets (*Nigerian Faces* and *ND Faces*) per face detector.

scores, and in red are impostor (cross-subject) scores. In Experiment 3, we remarked a larger average distance for cross-race impostor image pairs. Initially, we supposed this could be evidence of a discrepancy in biometric performance across races (at least with the ArcFace model). However, upon comparing the results from Experiment 3 to those of Experiment 4, we observed a similar mean location for the same-race impostor distributions, especially in the Caucasian/Asian face images from *ND Faces*. Such a similarity in the three impostor distributions suggested that the differences in performance may not be solely due to racial differences.

An interesting counterargument to this claim stems from the discrepancy in d' scores between the genuine and impostor distributions for each race. For the *Nigerian Faces* data, the d' value between genuine and impostor distributions was approximately 5.586; the d' value for the *ND Faces* data was approximately 6.68. This difference suggests that ArcFace is more capable of properly discriminating Caucasian/Asian face images than African face images.

The larger d' value (from *ND Faces*) is also visually observable in the narrower shape of the genuine (green diamonds) and impostor (thin red x markers) distributions. The distributions boast smaller standard deviations of 0.1237 and 0.1187, respectively. On the other hand, the thicker genuine (green squares) and impostor (thick red x markers) distributions from the *Nigerian Faces* data maintain larger standard deviations of 0.1560 and 0.1223, also respectively. This finding suggests that same-subject face images in the *Nigerian Faces* data set are more spread out in the ArcFace latent space whereas same-subject face images in the *ND Faces* data set are much more compact. Based on these results, there exists an opportunity and a possible need for re-training ArcFace with an increased number of African face images during the training process.

5.5. Face Detection Yield Results

In a smaller set of tangential experiments, we evaluated four face detectors' performance on all face images in *Nigerian Faces* and *ND Faces*. The chosen detectors reflect a variety of techniques used throughout the history of face (and general object) detection, starting with handcrafted Haar cascades, which do not employ deep learning methods, but perform near real-time without GPU usage. More recently developed are dlib and RetinaFace, which rely on convolutional neural network-based architectures to achieve state-of-the-art detection accuracy; MTCNN uses a cascade of CNN's to achieve this same degree of accuracy. As seen in Table 3, nearly all of the detectors showed face detection yields of greater than 0.97: six of them greater than 0.99. Among the four detectors, the Haar cascade method was the technique that showed the largest discrepancy between the African and Caucasian/Asian data sets. Similar to the previous suggestion to re-train ArcFace, we propose re-training Haar cascades with a new data set that includes more African face images. However, given the near perfect performance of the advanced detectors, we cannot claim there exists a substantial difference in face detection yield between Caucasian/Asian and African face images.

6. Conclusion

The influence of 28 African (Nigerian) ethnicities on face identification performance in relation to a Caucasian/Asian face set was studied in this paper. The effects were analysed from the performance metrics while focusing on the genuine and impostor score distributions. The results in this paper confirm that some of the Nigerian ethnicities have characteristic facial features that differentiate them from others, but however, they do not significantly affect biometric performance metrics. Results show that performance of face identification of varying Nigerian ethnicities does not depend on the characteristic facial features of the ethnicities involved. Additionally, this paper analyzes a partner Caucasian/Asian data set (*ND Faces*) for evidence of racial bias in face detection or recognition. Re-training of a select few models would provide useful to the community both detection and recognition tasks and may be endeavoured in future work.

References

[1] Salem Hamed Abdurrahim, Salina Abdul Samad, and Aqilah Baseri Huddin. Review on the effects of age, gender, and race demographics on automatic face recognition, nov 2018.

[2] Ajayi Adegboyega and Omojeje Ajishola. Oyo-Ondo relations: A study in pristine inter-group relations in Nigeria. *African Journal of History and Culture*, 11(6):57–64, nov 2019.

[3] Akinlolu Abdulazeez Adelaja, Atoyebi Michael, Akinola Oluwole Busayo, Fatunke Ibukun, Jimoh Rabiu, and Ajao Moyosore Salihu. Sex and Tribal differences in Facial measurements of Nigerians of Yoruba, Hausa and Igbo origin. *Research Journal of Forensic Sciences*, 4(1):5–11, 2016.

[4] Moses Olusola Adetona and Mathew Temitayo Shokunbi. Quantitative anthropometric and dermatoglyphic variation of

the major ethnic populations in Nigeria. *Journal of Experimental and Clinical Anatomy*, 18:55–62, 2019.

[5] Adelaja Abdulazeez Akinlolu. Facial biometrics of Yorubas of Nigeria using Akinlolu-Raji image-processing algorithm. *Journal of Medical Sciences*, 36(2):45, mar 2016.

[6] I. Anas. Anthropometric Comparison of Nasal Indices between Hausa and Yoruba Ethnic Groups in Nigeria. *Journal of Scientific Research and Reports*, 3(3):437–444, jan 2014.

[7] Elvin Bhaskar and Chung H. Kau. A Comparison of 3D Facial Features in a Population from Zimbabwe and United States. *European Journal of Dentistry*, 14(1):100–106, feb 2020.

[8] Stefan Böhringer and Markus A de Jong. Quantification of facial traits. *Frontiers in genetics*, 10:397, 2019.

[9] Kevin Bowyer and Michael King. Why face recognition accuracy varies due to race. *Biometric Technology Today*, 2019(8):8–11, sep 2019.

[10] Jacqueline G. Cavazos, Eilidh Noyes, and Alice J. O'Toole. Learning context and the other-race effect: Strategies for improving face recognition. *Vision Research*, 157:169–183, apr 2019.

[11] Jacqueline G. Cavazos, P. Jonathon Phillips, Carlos D. Castillo, and Alice J. O'Toole. Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *Arxiv.org*, pages 1–11, dec 2019.

[12] Joanne Burnette Cole. *Heritability and Genetic Determinants of Normal Human Facial Shape and Body Size*. PhD thesis, University of Maryland, Maryland, 2016.

[13] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2019-June, pages 4685–4694. IEEE Computer Society, jun 2019.

[14] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[15] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*, 2019.

[16] T Dunstone and N. Yager. System Evaluation: The Statistical Basis of Biometric Systems. In *Biometric System and Data Analysis*, pages 111–152. Springer US, oct 2008.

[17] O Ebeye, E Emore, E Ebite, and N Ijeh. Facial Dimensions In Urhobo's Of Nigeria. *The Internet Journal of Biological Anthropolog*, 4(1):1–4, 2009.

[18] D. E. O. Eboh and E. A. John. Morphological Assessment of face and Nose Shapes among the Ukwuanis of Delta State, Nigeria. Technical report, Delta State University, Abraka, 2010.

[19] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Rodrigo Ros-Gomez, and Judith Liu-Jimenez. Small fingerprint scanners used in mobile devices: The impact on biometric performance. *IET Biometrics*, 5(1):28–36, mar 2016.

[20] Macarena Fuentes-Guajardo. *Identification of genes responsible for the variation in facial and teeth morphology in Latin Americans*. PhD thesis, University College London, London, dec 2019.

[21] John J. Howard, Yevgeniy B. Sirotin, and Arun R. Vemury. The Effect of Broad and Specific Demographic Homogeneity on the Imposter Distributions and False Match Rates in Face Recognition Algorithm Performance. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, sep 2019.

[22] Katia Huri, Eli Omid David, and Nathan S. Netanyahu. DeepEthnic: Multi-label ethnic classification from face images. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11141 LNCS, pages 604–612. Springer Verlag, oct 2018.

[23] Herbert Igboanusi and Hans Georg Wolf. The role of ethnically mixed marriages in language shift: A case study of Nigeria's minority languages. *Sociolinguistic Studies*, 3(3):451–464, jun 2009.

[24] International Organization for Standardization. ISO/IEC 19795-1:2006: Information technology — Biometric performance testing and reporting — Part 1: Principles and framework. Technical report, 2006.

[25] Rabiu O. Jimoh, Sulyman B. Alabi, Aremu Shuaib Kayode, Ajao M. Salihu, and Olubukola D. Ogidi. Rhinometry: Spectrum of nasal profile among Nigerian Africans. *Brazilian Journal of Otorhinolaryngology*, 77(5):589–593, sep 2011.

[26] Hyun-Jin Kim, Sun-Wha Im, Ganchimeg Jargal, Siwoo Lee, Jae-Hyuk Yi, Jeong-Yeon Park, Joohon Sung, Sung-Il Cho, Jong-Yeol Kim, Jong-Il Kim, and Jeong-Sun Seo. Heritabilities of Facial Measurements and Their Latent Factors in Korean Families. *Genomics & Informatics*, 11(2):92, 2013.

[27] Davis E King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.

[28] K S Krishnapriya, V Albiero, K Vangara, M C King, and K W Bowyer. Issues Related to Face Recognition Accuracy Varying Based on Race and Skin Tone. *IEEE Transactions on Technology and Society*, 1(1):8–20, 2020.

[29] K. S. Krishnapriya, Kushal Vangara, Michael C. King, Vitor Albiero, and Kevin Bowyer. Characterizing the variability in face recognition accuracy relative to race. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2019-June, pages 2278–2285, Long Beach, CA, jun 2019. IEEE Computer Society.

[30] Yi Li, Wenting Zhao, Dan Li, Xianming Tao, Ziyi Xiong, Jing Liu, Wei Zhang, Anquan Ji, Kun Tang, Fan Liu, and Caixia Li. EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population. *Human Genetics*, 138(6):681–689, jun 2019.

[31] Neil A. Macmillan and Douglas C. Creelman. *Detection Theory: A User's Guide*. Taylor & Francis, 2 edition, 2004.

[32] D. R Omotoso, O. O Oludiran, and C. L Sakpa. Nasofacial Anthropometry of Adult Bini Tribe In Nigeria. *African Journal of Biomedical Research*, 14(3):219–221, 2011.

[33] Margaret Peil. Interethnic Contacts in Nigerian Cities. *Africa: Journal of the International African Institute*, 45(2):107–122, 1975.

[34] Victoria V. Rostovtseva, Anna A. Mezentseva, Sonja Windhager, and Marina L. Butovskaya. Second-to-fourth digit ratio and facial shape in Buryats of Southern Siberia. *Early Human Development*, 149:105138, oct 2020.

[35] Usman Tariq, Yuxiao Hu, and Thomas S. Huang. Gender and ethnicity identification from silhouetted face profiles. In *Proceedings - International Conference on Image Processing, ICIP*, pages 2441–2444, Cairo, Egypt, nov 2009. IEEE Computer Society.

[36] M. G. Taura, L. H. Adamu, and A. Gudaji. Variation of facial features among three African populations: Body height match analyses. *HOMO- Journal of Comparative Human Biology*, 68(1):69–79, jan 2017.

[37] Kpela Terkula, S. S. Adebisi, and B. Danborno. Facial Anthropometry of Adult Tiv and Idoma Tribes of Nigeria. *Global Advanced Research Journal of Medicine and Medical Sciences (GARJMMS)*, 8(2):009–013, 2019.

[38] Olukanni A. Titilayo, Amoo O. Samuel, Olukanni O. David, and Taiwo I. Adewunmi. Genetic variations among three major ethnic groups in Nigeria using RAPD. *Molecular Biology Research Communications*, 7(2):51–58, 2018.

[39] A. J. N. Tremearne. Notes on Some Nigerian Tribal Marks. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 41(1):178, jan 1911.

[40] Dimosthenis Tsagkrasoulis, Pirro Hysi, Tim Spector, and Giovanni Montana. Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific Reports*, 7(1):1–18, apr 2017.

[41] Ukoha Ukoha Ukoha, Onochie Okwudili Udemezue, Chidi Kingsley Oranusi, Azuoma Lasbrey Asomugha, Uchechukwu Dimkpa, and Lynda Chinenye Nzeukwu. Photometric facial analysis of the Igbo Nigerian adult male. *Nigerian medical journal : journal of the Nigeria Medical Association*, 53(4):240–244, oct 2012.

[42] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[43] Seth M. Weinberg, Jasmien Roosenboom, John R. Shaffer, Mark D. Shriver, Joanna Wysocka, and Peter Claes. Hunting for genes that shape human faces: Initial successes and challenges for the future. *Orthodontics & Craniofacial Research*, 22(S1):207–212, may 2019.

[44] Wei Wu, Guiying Zhai, Zejun Xu, Bo Hou, Dahua Liu, Tianyi Liu, Wei Liu, and Fu Ren. Whole-exome sequencing identified four loci influencing craniofacial morphology in northern Han Chinese. *Human Genetics*, 138(6):601–611, jun 2019.

[45] Zhaoming Wu and Han Sung Jung. How the diversity of the faces arises, dec 2019.

[46] Jia Xiang and Gengming Zhu. Joint face detection and facial expression recognition with mtcnn. In *2017 4th international conference on information science and control engineering (ICISCE)*, pages 424–427. IEEE, 2017.

[47] Ziyi Xiong, Gabriela Dankova, Laurence J. Howe, Myoung Keun Lee, Pirro G. Hysi, Markus A. De Jong, Gu Zhu, Kaustubh Adhikar, Dan Li, Yi Li, Bo Pan, Eleanor Feingold, Mary L. Marazita, John R. Shaffer, Kerrie McAloney, Shuhua Xu, Li Jin, Sijia Wang, Femke M. De Vri, Bas Lendemeije, Stephen Richmond, Alexei Zhurov, Sarah Lewis, Gemma Sharp, Lavinia Paternoster, Holly Thompson, Rolando Gonzalez-Jose, Maria Catira Bortolini, Samuel Canizales-Quinteros, Carla Gallo, Giovanni Poletti, Gabriel Bedoya, Francisco Rothhammer, André G. Uitterlinden, M. Arfan Ikram, Eppo B. Wolvius, Steven A. Kushner, Tamar Nijsten, Robert Jan Palstra, Stefan Boehringer, Sarah E. Medland, Kun Tang, Andrés Ruiz-Linares, Nicholas G. Martin, Timothy D. Spector, Evie Stergiakouli, Seth M. Weinberg, Fan Liu, and Manfred Kayser. Novel genetic loci affecting facial shape variation in humans. *eLife*, 8:1–11, nov 2019.