Measuring Hidden Bias within Face Recognition via Racial Phenotypes

Seyma Yucer, Furkan Tektas, Noura Al Moubayed and Toby P. Breckon

Durham University

Motivation

Ambiguous Definition of Race: The historical and biological definitions of race vary and racial context is not fixed over time [1].

Privacy of Protected Attributes: Exposing demographic origin with in face recognition studies may identify the representation of a particular group, leading to the potential for racial profiling and associated targeting [2].

Confined Groupings: Skin or racial grouping strategies limits the scope of any study as they fail to capture the whole aspect of the racial bias problem within face recognition where it needs to consider both multi-racial or less stereotypical members of such groups.

Racial Appearance Bias: Studies [3,4] show that individuals with more stereotypical racial appearance suffer poorer outcomes than those with less stereotypical appearance for their race. A better understanding of the role of phenotypic variation complements solutions for both racial and racial appearence bias.

> **This study introduces an alternative racial bias analysis methodology via facial phenotype attributes for face recognition.**

References

[1] Jayne Chong-Soon Lee. Navigating the topology of race,1994

[2] Paul Mozur. One month, 500,000 face scans: How china is using AI to profile a minority.The New York Times, 2019

[3] Keith B Maddox and Jennifer M Perry. Racial appearance bias: Improving evidence-based policies to address racial disparities. Policy Insights from the Behavioral and Brain Sciences, 2018.

[4] Allison L Skinner and Gandalf Nicolas. Looking black or looking back? using phenotype and ancestry to make racial categorizations. Journal of Experimental Social Psychology, 2015

For more information

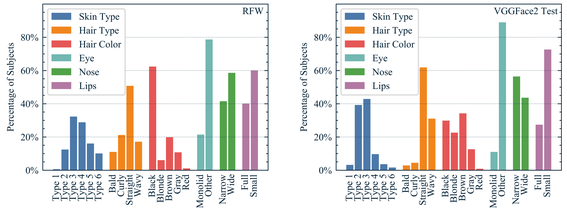
arXiv

GitHub

Racial Phenotypes for Face Recognition

We propose using race-related facial (phenotype) characteristics within face recognition to investigate racial bias by categorising representative racial characteristics on the face and exploring the impact of each characteristic phenotype attribute: **skin types, eyelid type, nose shape, lips shape, hair colour and hair type.**

The distribution of racial phenotype attributes



We annotate the phenotype attributes on each subjects of RFW and VGGFace2 benchmark datasets. For both datasets, we **observe that the dominant phenotype attribute categories are Skin Type 2/3, Straight Hair, Narrow Nose, Other (non-monolid) Eyes, Small Lips, which correlates to the dominant presence of Caucasian faces** as can be seen on the figure above.

Results

Subgroup-based face verification performance

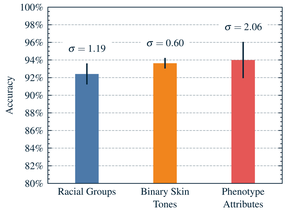
On RFW, sorted by descending order of accuracy

Skin	Lips	Eye	Nose	Hair Type	Ratio (%)	Accuracy (%)	Skin	Lips	Eye	Nose	Hair Type	Ratio (%)	Accuracy (%)
{1,2}	Small	Other	Narrow	Straight	3.82	96.53	{3,4}	Full	Monolid	Wide	Straight	1.55	91.63
{3,4}	Small	Other	Narrow	Straight	7.43	96.45	{1,2}	Small	Other	Narrow	Bald	0.28	91.29
{3,4}	Small	Other	Narrow	Wavy	3.67	96.11	{5,6}	Full	Other	Narrow	Curly	1.97	91.23
{1,2}	Small	Other	Wide	Straight	3.03	95.63	{3,4}	Small	Other	Wide	Bald	1.68	91.01
{1,2}	Small	Other	Narrow	Wavy	1.64	95.62	{1,2}	Full	Other	Narrow	Wavy	0.27	90.74
{1,2}	Full	Other	Narrow	Straight	0.70	95.59	{3,4}	Small	Monolid	Wide	Wavy	0.96	90.17
{3,4}	Full	Other	Narrow	Straight	3.59	95.28	{1,2}	Small	Other	Wide	Bald	0.46	89.78
{3,4}	Full	Other	Wide	Straight	4.47	94.98	{5,6}	Small	Other	Narrow	Curly	0.81	89.50
{3,4}	Small	Other	Wide	Wavy	2.95	94.92	{3,4}	Small	Monolid	Narrow	Wavy	1.20	89.35
{3,4}	Small	Other	Wide	Straight	8.83	94.92	{5,6}	Full	Other	Wide	Curly	13.09	89.18
{1,2}	Full	Other	Wide	Straight	0.33	94.87	{3,4}	Full	Other	Wide	Bald	0.80	86.02
{1,2}	Small	Other	Wide	Wavy	0.72	94.56	{5,6}	Small	Other	Wide	Bald	0.99	85.90
{3,4}	Small	Other	Wide	Curly	0.51	93.89	{3,4}	Full	Other	Wide	Curly	0.46	85.38
{3,4}	Full	Other	Wide	Wavy	1.90	93.41	{3,4}	Small	Monolid	Narrow	Bald	0.32	84.10
{3,4}	Full	Other	Narrow	Wavy	1.94	93.10	{5,6}	Small	Other	Narrow	Bald	0.30	82.81
{3,4}	Small	Other	Narrow	Bald	0.68	92.50	{3,4}	Small	Monolid	Wide	Bald	0.52	82.67
{3,4}	Small	Other	Narrow	Curly	0.31	92.45	{3,4}	Full	Monolid	Narrow	Wavy	0.43	82.04
{5,6}	Small	Other	Wide	Curly	2.81	92.23	{5,6}	Full	Other	Narrow	Bald	0.53	81.24
{3,4}	Small	Monolid	Wide	Straight	6.59	91.93	{1,2}	Small	Monolid	Narrow	Straight	0.47	81.04
{3,4}	Full	Monolid	Narrow	Straight	1.81	91.78	{3,4}	Full	Monolid	Wide	Wavy	0.27	79.47
{5,6}	Full	Other	Wide	Bald	3.62	91.74	{5,6}	Full	Other	Wide	Wavy	0.32	78.94
{3,4}	Small	Monolid	Narrow	Straight	7.95	91.70							
σ													5.07

We create various subgroups where each subgroup has same phenotypic attribute combinations. Our main purpose of such pairing is to show ***what would change when only one attribute changes, but other attributes remain the same?***
Groups who have one of the **attributes like wide nose, full lips, and monolid eye type always have less accuracy than the other groups with a narrow nose, small lips and other eye** (when rest of the attributes are same).

Accuracy variations for three grouping strategies

Standard deviation of the groupings reflects the amount of measured bias.

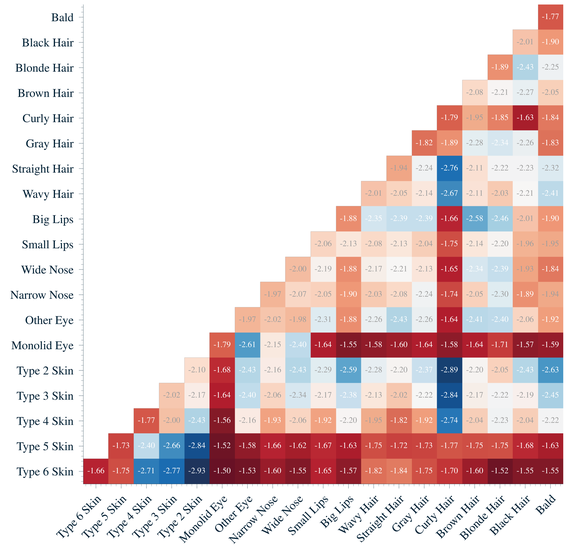


We compare racial groupings *{African, Asian, Indian, Caucasian}*, **binary skin tone groupings** *{lighter skin-tone, darker skin-tone}* **strategies with our phenotype-based strategy.**

We show that accuracy andstandard deviation differs across all three strategies. Higher variation reveals hidden bias, which may be missed in narrow, erroneous racial or binary skin tones grouping strategies. **The phenotype-based grouping strategy brings a more granular observation of the variability in performance and hence a more resolute measure of performance bias.**

Cross-attribute based pairings false matching rate

Each cell depicts FMR on a logarithmic scale which is log10(FMR) with lower negative values (close to zero) encoding superior false match rates.



Above we pair each attribute category with all other attribute categories to assess cross-attribute pairing performancen - *we clearly show that Type 5, Type 6 and monolid eyes pairings have higher false positive matching rates than others.*

Conclusion

In this study, **we demonstrate that phenotype-based evaluation strategy reveals racial bias comprehensively whilst avoiding exposing potentially protected or ill-defined attributes.**

Furthermore, **we observe apparent performance differences between race-welated phenotype attribute categories and subgroups**.

Future work will focus on improving facial appearance variations to provide more balanced and realistic test scenario distributions.