

# SREDS: A dichromatic separation based measure of skin color

Keivan Bahmani

Clarkson University

Potsdam, NY, USA

bahmank@clarkson.edu

Richard Plesh

Clarkson University

Potsdam, NY, USA

pleshro@clarkson.edu

Chinmay Sahu

Clarkson University

Potsdam, NY, USA

sahuc@clarkson.edu

Mahesh Banavar

Clarkson University

Potsdam, NY, USA

mbanavar@clarkson.edu

Stephanie Schuckers

Clarkson University

Potsdam, NY, USA

sschucke@clarkson.edu

**Abstract**—Face recognition (FR) systems are fast becoming ubiquitous. However, differential performance among certain demographics was identified in several widely used FR models. The skin tone of the subject is an important factor in addressing the differential performance. Previous work has used modeling methods to propose skin tone measures of subjects across different illuminations or utilized subjective labels of skin color and demographic information. However, such models heavily rely on consistent background and lighting for calibration, or utilize labeled datasets, which are time-consuming to generate or are unavailable. In this work, we have developed a novel and data-driven skin color measure capable of accurately representing subjects' skin tone from a single image, without requiring a consistent background or illumination. Our measure leverages the dichromatic reflection model in RGB space to decompose skin patches into diffuse and specular bases.

**Index Terms**—face recognition, differential performance, skin reflectance, skin color

## I. INTRODUCTION

Interest in facial recognition has been increasing rapidly as the technology has improved in performance and reliability over the past few decades. Facial recognition systems are commonly used in video authentication, criminal identification and building/device access control and many other areas [1]. Since facial recognition is involved in such critical applications, researchers are investigating how error rates differ between different demographic groups. A report by the National Institute of Standards and Technology (NIST) studied this question and found evidence of demographic differentials in the majority of algorithms evaluated [2]. While the best algorithms did not present a differential, there is a desire to minimize this for all algorithms. Performance differential commonly comes about by maximizing overall predictive accuracy without considering how one subgroup's performance masks another's deficiencies [3]. Recent research has focused on providing solutions to differential performance in facial recognition, or often called "bias" by the popular press. The problem has two components, false negatives and false positives. Since each is a unique problem and bias encompasses a larger, diverse set of issues, biometric researchers prefer the term differential performance

rather than bias [4]. The solutions typically fall into one of three categories: re-balanced training sets [5], protected attribute suppression [6], and model adaption [7]. Nearly all of these solutions require datasets containing demographic labels for applying their strategies and evaluating performance. While datasets like this exist [8], [9], they represent a small fraction of the total number of datasets for facial recognition. Furthermore, datasets are often labeled for a single task [10], and demographic data may be overlooked because it is difficult to collect reliably. For this reason, many researchers apply demographic information to datasets after collection. One approach is to use an off-the-shelf deep-learning-based ethnicity classifier. The authors in [5] utilize the proprietary Face++ API for labeling their datasets. While this method scales easily to large datasets, the failure points of these models are not well understood and the complex relationships they rely on are very difficult to interpret. Due to this limitation, many researchers rely on more readily apparent attributes such as skin color for labeling their datasets. While skin color doesn't directly represent demographic information, it does have correlation with ethnic self-definitions [11]. Nevertheless, determining a person's inherent skin color from an image can be challenging due to a person's natural skin variability, the camera parameters, and changes in lighting. Figure 1 depicts the variation of skin tone for a subject in MEDS-II dataset [8]. A widespread method for estimating skin pigment from facial images is via Fitzpatrick skin type (FST) [12]. Despite its popularity, evidence is beginning to mount that FST has limited quantification and reliability [13], particularly for non-white individuals [14]. In response to FST's limitations and off-the-self deep algorithm's poor interpretability, researchers have developed the skin color metrics Individual Typology Angle (ITA) [15] and Relative Skin Reflectance (RSR) [16]. ITA utilizes colorimetric parameters to provide a point-wise estimate of the skin color represented in the image of a person. On the other hand, RSR is a data-driven approach that utilizes the distribution of skin pixels in color space to fit a linear model estimating skin tone. Both methods are sensitive to changes in illumination and RSR in particular requires a highly controlled acquisition environment (constant background, lighting and camera). These restrictions greatly hamper the usefulness of these metrics in the more challenging

This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation under Grant No.650503.

deployments of facial recognition without control over the collection environment.

Motivated by these shortcomings, we introduce Skin Reflectance Estimate based on Dichromatic Separation (SREDS). SREDS provides a continuous skin tone estimate by leveraging the dichromatic reflection model [17] and explicitly considering the different types of illumination across the face. This provides SREDS with superior or comparable performance in both less consistent and highly controlled acquisition environments. Additionally, the dichromatic model provides us greater interpretability into the locations of the face most utilized in the final metric generation.

To evaluate this measure, we consider both its stability and meaningfulness compared to ITA and RSR over three different datasets: Multi-PIE Multi-view dataset [18], the Multiple Encounter Dataset (MEDS-II) [8], and the Morph dataset [9]. We measure consistency using intra-subject variation and analyze meaningfulness by examining the distribution of ethnicities with respect to skin color estimates. Since bias mitigation frameworks optimize to reduce differential performance with respect to their demographic information, we expect any performance improvement in the demographic labeling to translate to more fair algorithms for end users. The next sections provide a description of skin metrics, preprocessing, datasets used, results, discussion, and conclusions.

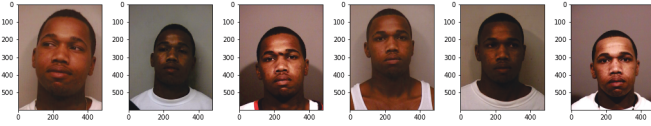


Fig. 1. Variability of skin color of the same subject from MEDS-II dataset

## II. BENCHMARK SKIN COLOR METRICS

In this section we discuss our implementation of two existing skin color metrics, ITA and RSR. Each method has a unique pipeline from facial image to final metric, and we replicated the original work to the best of our ability.

### A. Individual Typology Angle

Individual typology angle (ITA) is a type of colorimetric analysis designed to measure acquired tanning [15]. Given the simplicity of ITA and its correlation with Melanin Index [19], it is an ideal candidate for determining skin tone directly from an image. An RGB image is converted into CIE-Lab space [20], as follows: (1) the ‘L’ component which quantifies luminance, (2) the ‘a’ component - absence or presence of redness, and (3) the ‘b’ component - yellowness. Using the ‘L’ and ‘b’ components, Pixel-wise ITA value, in degrees, can be estimated throughout an image as:

$$ITA = \frac{\arctan(L - 50)}{b} * \frac{180}{\pi}. \quad (1)$$

In order to find suitable skin pixels in the image, a Dlib landmark extractor (see Section V) is used to detect the forehead, left cheek, and right cheek facial regions. For

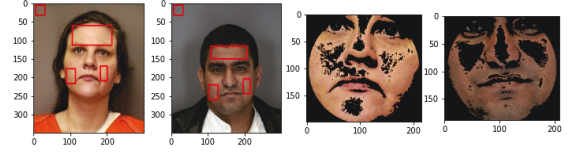


Fig. 2. Region of interest extraction for cheeks, forehead, and background crops (Left). Adaptive skin segmentation (Right)

each facial region, pixel-wise ITA is computed and smoothed using an averaging filter. The mode of each region’s resulting distribution are averaged together to create a single skin tone estimate for a face.

### B. Relative Skin Reflectance

Relative Skin Reflectance (RSR) is a process designed to relate physical properties of the skin to the performance of facial recognition [16]. The pipeline works by removing the confounding effects of imaging artifacts on skin pixels and then fitting a line in the direction of greatest variance in RGB color-space. The resulting metric is related to the skin tone of each subject relative to the rest of the photos in the dataset. The computation of RSR begins with the selection of facial skin pixels using face detection, circular masking, and luminance outlier removal as described in [21]. Pixel intensities are then corrected using divisive normalization via background sampling. We refer to this preprocessing pipeline as *adaptive skin segmentation*. Figure 2 depicts the outcome of adaptive skin segmentation algorithm applied to two faces. After removal of non-skin and outlier pixels, the researchers assume that the direction of greatest variation in RGB color-space represents net skin reflectance. Linear principal component analysis (PCA) is used to fit a line in this direction.

The final RSR metric is produced by averaging the projection of skin pixels onto the first principal component. Assumptions include consistent lighting, the same acquisition camera, and constant background. The Multi-PIE dataset was able to meet all conditions. However, due to the lack of constant background in MEDS-II and MORPH-II, the background normalization step was bypassed for these datasets. As a further limitation, the metric only gives an indication of where a subject lies in terms of net skin reflectance relative to the other subjects in the dataset, rather than an absolute measure.

## III. HUMAN SKIN ANALYSIS

Previous work relied on heavily controlled and calibrated image acquisition, *i.e.*, acquiring multiple samples under controlled illumination or with different polarization, in order to study the fundamentals of reflectance. The acquired data is fit to bidirectional reflectance distribution function (BRDF) and bidirectional surface scattering distribution function (BSSRDF) models [22] to fully represent the spatially-varying reflectance of the human face [23], [24]. We leverage the dichromatic reflection model [17] for our proposed

“Skin Reflectance Estimate based on Dichromatic Separation” (SREDS) metric. SREDS generates data-driven, additive, and interpretable measure of skin color.

The Dichromatic Reflection Model (DRM) is a general model for estimating the intrinsic reflections from a standard RGB image. DRM defines two types of reflections, interface (*specular*) and body (*diffuse*), where each reflection component can be further decomposed into spectral distributions and geometric scaling factors [17]. The general reflection model proposed in DRM is given in two forms:

$$L(\lambda, i, e, g) = L_i(\lambda, i, e, g) + L_b(\lambda, i, e, g) \quad (2)$$

$$= m_i(i, e, g)c_i(\lambda) + m_b(i, e, g)c_b(\lambda), \quad (3)$$

where in Equation 2,  $L_i$ ,  $L_b$ ,  $\lambda$ ,  $i$ ,  $e$ , and  $g$ , respectively, represent radiance of interface reflection, radiance of body reflection, wavelength of light, angle of incidence, angle of existence, and phase angle. The factors  $L_i$  and  $L_b$  are further decomposed into *compositions*,  $(c_i, c_b)$ , and *magnitudes*,  $(m_i, m_b)$  as shown in Equation 3. Composition components (representing shading) depend on the geometry of the object and are independent of the wavelength. Magnitude components represent the wavelength dependent nature of the radiance (representing color) and are independent of the geometry of the object.

The DRM introduces a set of assumptions to further simplify Equation 3 and represent the pixel values in the RGB domain as a tristimulus integration over the amount of received light and sensitivity of the camera at each wavelength. The model assumes that the material of interest is in-homogeneous, opaque, and uniformly colored. Additionally, the material should have one significant specular reflection, isotropic diffuse reflection, and constant spectral distribution across the scene. The proposed tristimulus integration is a linear transformation. As a result, every pixel value in the image can be approximated by a linear combination of interface and body reflection colors of the material as

$$C_L = m_i C_i + m_b C_b, \quad (4)$$

where  $C_L$ ,  $C_i$ ,  $C_b$ ,  $m_i$ , and  $m_b$ , respectively, represent the pixel value, color of interface reflection, color of body reflection, magnitude of interface reflection, and magnitude of body reflection.

The specular reflection is assumed to have the same spectral power distribution (color) as the incident illumination. The amount of specular reflection relates to the angle of incidence, index of refraction of the material and polarization of the incoming illumination as governed by Fresnel’s laws [17], and is largely due to the air-oil interface from the surface of the skin [24], [25]. The diffuse component is generally assumed to be isotropic [17], and it is due to the subsurface scattering of light with melanin and hemoglobin components in epidermis and dermis layers of skin [24], [26].

#### IV. SKIN REFLECTANCE ESTIMATE BASED ON DICHROMATIC SEPARATION (SREDS)

In this paper, we introduce a new measure of skin color, called the Skin Reflectance Estimate based on Dichromatic Separation (SREDS). SREDS aims to decompose the light reflected off the skin into specular and diffuse components and construct a data-driven skin color metric from the diffuse component of the skin. The diffuse component is due to the interaction of the light with hemoglobin and melanin components in the lower layers of the skin. As a result, it is assumed to be *Lambertian*, *i.e.*, is reflected equally in all directions. Utilizing the diffuse component of the skin allows the SREDS to be insensitive to naturally accruing specular reflections due to uncontrolled illumination and be independent of the angle between the illumination source and the viewer (camera), thus circumventing the requirement for consistent illumination and background as needed to calculate RSR.

SREDS extracts patches of skin from the forehead, and right and left cheeks of each face image. Selecting smooth patches of skin allows us to minimize the effect of geometry on the estimated reflection components. Previous work on skin reflectance of human face suggests that cheeks and forehead have similar translucency and isotropic diffuse reflection [27]. This allow us to rely on the DRM to estimate the diffuse and specular components of the selected skin patches. Given the independence assumption between specular and diffuse components, lack of ground truth, and the smoothness of the selected skin patches (no need to account for the geometry of the face), our problem can be classified as a blind source separation (BSS) problem [28]. Independent component analysis (ICA) and principal component analysis (PCA) are two common BSS algorithms [28]. Previous work showed the effectiveness of ICA in decomposing melanin and hemoglobin components of skin on samples acquired using high quality cameras in a lab environment [29].

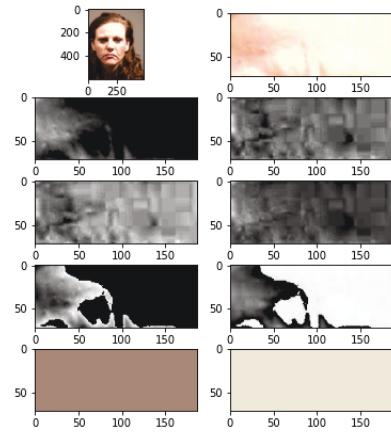


Fig. 3. Example of a skin patch decomposition. Skin patch of (forehead) (Row-1), PCA (Row-2), ICA (Row-3), and NNMF (Row-4) decompositions represented as grayscale images (black represents 0 and white represents 1), SREDS Bases (Row-5) - Specular (Right), Diffuse (Left).

We observed that in a few cases PCA resulted in meaningful but very noisy decomposition. However, ICA completely failed to produce a meaningful decomposition. Figure 3 represents the PCA (Row-2) and ICA (Row-3) decompositions for a patch of skin from subject’s forehead (Row-1). Note that, the sole BSS pipeline does not provide a method for identifying the diffuse component from the decomposed components. Given the ineffectiveness of PCA and ICA decompositions, we propose to relax the independent assumption and utilize the fact that our images and magnitudes ( $m_i$ ,  $m_b$ ) are all strictly positive. We propose to employ Non-negative Matrix Factorization (NMF) [30] to extract representations of specular and diffuse components from each patch of skin. NMF have shown to perform very well in image based BSS problems [31].

Additionally, unlike PCA, all the weights in NFM are strictly positive, *i.e.*, there is no subtraction involved in representing the input image [32]. This is in line with the combination of diffuse and specular components proposed in DRM and leads to better interpretability of the decomposed representations [32]. The NFM decomposition used in SREDS is given by

$$V^t \approx \sum_{i=1}^M w_i h^t = WH, \quad (5)$$

where  $V^t$  denotes the  $t$ th skin patch, represented as an  $n \times 3$  vector, where  $n$  denotes the number of pixels in each skin patch.  $W$  and  $H$  respectively represent our  $n \times 2$  and  $2 \times 3$  matrix factors. To find the  $W$  and  $H$  matrices, we solve the following constrained optimization problem:

$$\begin{aligned} & \underset{W, H}{\operatorname{argmin}} \quad \frac{1}{2} \|V^t - WH\|^2 \\ & \text{subject to } W \geq 0 \\ & \quad H \geq 0, \end{aligned} \quad (6)$$

where  $\|\cdot\|^2$  denotes the L2 norm. To satisfy the constraints, we employ non-negative double singular value decomposition (NNDSVD) with averaging initialization [30] to initialize  $W$  and  $H$  matrices when solving the optimization problem in Equation 5 numerically.

As the specular component reflects the color of the illumination and is generally brighter than the color of skin, we select the row of  $H$  with higher sum as our specular basis while the other row represents our diffuse basis. We observed that this simple assignment rule can be very effective over multiple datasets. The fourth and the fifth rows of Figure 3 respectively show the gray scale representation of our NFM decomposition and the RGB representation of the associated specular (right) and diffuse (left) bases. Finally, we utilize kernel principal component analysis (KPCA) [33] over our estimated diffuse bases and select the first principal component as our measure of skin color. We evaluated linear, polynomial, and radial basis function (RBF) kernels and observed that the degree-3 polynomial kernel worked best across the evaluated datasets. The final SREDS measure is an average of the three individual measures calculated from forehead and cheeks.

## V. REGION OF INTEREST SELECTION

In order to identify regions of interest (ROI) to compute the skin color metrics, we use facial landmark detection. Given an input image, a facial landmark predictor attempts to identify key points of interest based on the shape of the face using a two-step process. First, a face is detected in an image. Then, key facial landmarks are detected. In this paper, we use the Dlib library to detect key facial landmarks [34]. Dlib uses this pre-trained facial landmark detector to estimate the location of 68 coordinates that map to facial structures. Then using the landmarks near cheeks and the forehead, we extract rectangular patches from the right cheek, left cheek, forehead, and background for all subjects. The forehead patch is extracted by using the distance between the eye landmarks. The cheek landmarks extracted by estimating the distance between jawline and lips. An example of the extracted crops are shown in Figure 2.

## VI. DATASETS

1) *Multi-PIE*: The CMU Multi-PIE face database [18] contains more than 750,000 images of 337 people recorded in up to four sessions over five months. Subjects were imaged under 15 viewpoints and 19 illumination conditions while displaying a range of facial expressions. High-resolution frontal images were acquired as well. We selected three viewpoints (14\_0, 05\_1, 05\_0) where full views of the face were captured and images where the facial landmark detection failed are removed. Overall, we select 150,668 images out of 750,000 images from 314 subjects.

2) *MEDS-II*: Multiple Encounter Dataset (MEDS-II) [8] by NIST is a data corpus curated from a selection of deceased subjects with prior multiple encounters. It consists of 1170 images from 425 subjects. We utilized the 856 mugshot images for our research. Then, images where facial landmark detection failed are removed, resulting in a reduced image set consisting of 836 images from 171 subjects.

3) *Morph-II*: The academic MORPH database is a non-commercial dataset collected over 5 years with multiple images of each subject (longitudinal). It is not a controlled collection (*i.e.*, it was curated in real-world conditions). This dataset contains 55,139 unique images of more than 13,000 individuals, spanning from 2003 to late 2007. Ages of the subjects range from 16 to 77, with a median age of 33. The average number of images per individual is 4. The average time between photos is 164 days. Images where the facial landmark detection failed are removed, resulting in a reduced image set consisting of 55,063 images from 13,152 subjects.

## VII. RESULTS AND DISCUSSION

In this work, we evaluate and compare the performance of ITA, RSR, and SREDS under both controlled (Multi-PIE [18]) and uncontrolled illumination (MEDS-II [8], Morph-II [9]). There is no objective measure of ground truth without a direct measure of skin reflectance controlling for all factors. However, we propose that the hypothetical perfect skin color metric has to be stable, *i.e.* robust against all possible variations

TABLE I

INTRA-SUBJECT VARIABILITY OVER MULTIPLE DATASETS. S1: NUMBER OF SUBJECTS, S2: AVERAGE NUMBER OF SAMPLES PER SUBJECT, S3: TOTAL NUMBER OF SAMPLES (AFTER CLEANING), ITA: INDIVIDUAL TYPOLOGY ANGLE, RSR: RELATIVE SKIN REFLECTANCE, SREDS: SKIN REFLECTANCE ESTIMATE BASED ON DICHROMATIC SEPARATION, RSR\*: RELATIVE SKIN REFLECTANCE USING SKIN PATCHES INSTEAD OF ADAPTIVE SEGMENTATION.

| Dataset                     | ITA          | RSR   | SREDS        | RSR*  | # S1   | # S2 | # S3    |
|-----------------------------|--------------|-------|--------------|-------|--------|------|---------|
| Multi-PIE - High Resolution | 0.401        | 0.307 | <b>0.138</b> | 0.356 | 314    | 3.6  | 1,170   |
| Multi-PIE - Multi View      | 0.926        | 0.860 | <b>0.820</b> | 0.881 | 314    | 447  | 150,668 |
| MEDS-II                     | <b>0.448</b> | 0.493 | 0.463        | 0.475 | 171    | 3.3  | 834     |
| Morph-II                    | 0.645        | 0.539 | <b>0.419</b> | 0.562 | 13,152 | 4.1  | 55,045  |

such as changes in illumination or camera sensitivity; thus, it is highly desirable for any skin color metric to have lower intra-subject variability. In order to fairly compare the evaluated metrics, we individually normalize each metric to have zero mean and unit variance and use the intra-subject variability (standard deviation) of each metric to evaluate its effectiveness. Table I presents the average intra-subject variability of the ITA, RSR, and SREDS over the evaluated datasets.

The image acquisition environment of Multi-PIE dataset provides us with both a constant background and controlled variation of illumination. This allow us to isolate the effect of illumination change on the intra-subject variability of the evaluated metrics. Our results suggests that SREDS outperforms the other two algorithms in both controlled and varying illumination environments and without relying on the constant background. Note that all measures show higher intra-subject variability under varying illumination. Figure 4 illustrates SREDS over multiple samples of the same subject collected while changing the angle incidence from  $-45^\circ$  (most left) to  $45^\circ$  (most right) with step size of  $15^\circ$ . SREDS is able to extract a stable diffuse component while the specular component is brighter with the increase of the angle of incident.

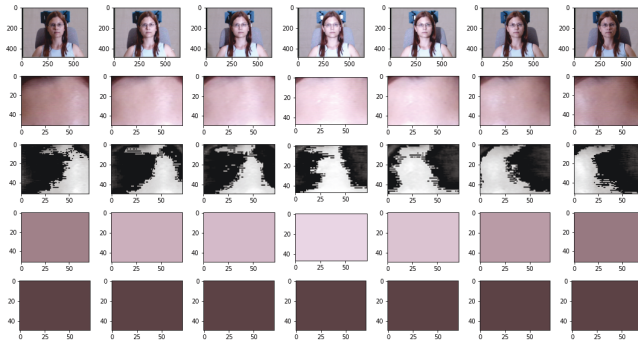


Fig. 4. Diffuse and specular components vs. illumination change in Multi-PIE dataset. Row-1: face images, Row-2: forehead skin patches, Row-3: weights associated with the specular basis, Row-4: specular bases, Row-5: diffuse bases. Best viewed in color.

We also evaluated all three algorithms using MEDS-II and Morph-II datasets (uncontrolled illumination). SREDS outperforms RSR and ITA in the larger and wider Morph-II dataset. However, we also observe that ITA provides marginally better performance than SREDS and RSR in the MEDS-II dataset. We suspect this might be due to the smaller size of the MEDS-II and the data-driven nature of the RSR and SREDS.

Additionally, we investigated the relationship between each measure and the ethnicity labels provided in the datasets. Figure 5 depicts the distribution of black and white subjects over each measure. The distribution of SREDS measurements over black and white subjects in uncontrolled illumination suggests that SREDS provides a meaningful progression from darker to lighter skin tone subjects without relying on consistent background, illumination or camera sensitivity.

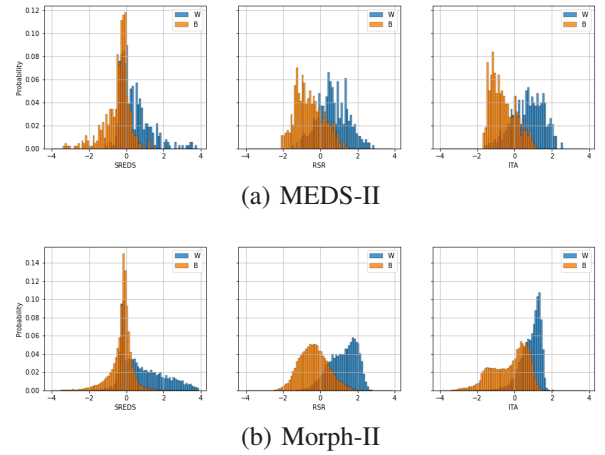


Fig. 5. Histograms of SREDS, RSR and ITA for black (B) and white (W) subjects in MEDS-II and Morph-II datasets. Best viewed in color.

#### A. Limitations and Future work

Our results suggests that SREDS can produce a continuous and interpretable skin color metric across different image acquisition environments and without relying on constant background or illumination. However, similar to the other evaluated measures, SREDS is still susceptible to illumination change. In future work we aim to improve upon this work by integrating a more complex basis assignment process. Our current model does not utilize the wavelength dependent nature of reflection in the assignment process and treats Red, Green, and Blue channels equally. Utilizing this fact can potentially improve the bases assignment process. Finally, in future work we plan to evaluate the correlation between the estimated SREDS values and direct measurements of skin reflectance using a DSM III skin colorimeter.

### VIII. CONCLUSIONS

While there is currently a large amount of active research providing solutions for removing bias from facial recognition systems, many of the methodologies assume the existence of large-scale data that is labeled and in-domain. Operational systems rarely can fulfil this requirement. These restrictions make the solutions very difficult to apply outside of the structured and labeled datasets generally used to benchmark algorithms. Inspired by this problem, we developed an interpretable skin tone estimate with few restrictions that can provide performance comparable or superior to similar methods in the literature. This estimate will provide a way for facial recognition applications without access to large demographically labeled datasets within their domain to make effective use of methods to reduce difference performance and promote fairness in face recognition.

### REFERENCES

- [1] Y. Kortli, M. Jridi, A. A. Falou, and M. Atri, "Face Recognition Systems: A Survey," *Sensors (Basel, Switzerland)*, vol. 20, no. 2, Jan. 2020.
- [2] P. J. Grother, M. L. Ngan, and K. K. Hanaoka, "Face Recognition Vendor Test Part 3: Demographic Effects," *NIST Pubs*, Dec. 2019.
- [3] D. Madras, E. Creager, T. Pitassi, and R. Zemel, "Learning Adversarially Fair and Transferable Representations," in *International Conference on Machine Learning*. PMLR, Jul. 2018, pp. 3384–3393, iSSN: 2640-3498. [Online]. Available: <http://proceedings.mlr.press/v80/madras18a.html>
- [4] J. J. Howard, Y. B. Sirotin, and A. 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)*, Sep. 2019, pp. 1–8, iSSN: 2474-9699.
- [5] M. Wang and W. Deng, "Mitigating Bias in Face Recognition Using Skewness-Aware Reinforcement Learning," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pages: 9328.
- [6] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning Agnostic Representations with Application to Face Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.
- [7] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial Faces in the Wild: Reducing Racial Bias by Information Maximization Adaptation Network," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 692–702, iSSN: 2380-7504.
- [8] A. P. Founds, N. Orlans, W. Genevieve, and C. I. Watson, "Nist special database 32-multiple encounter dataset ii (meds-ii)," NIST, Tech. Rep., 2011.
- [9] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE, 2006, pp. 341–345.
- [10] M. Alvi, A. Zisserman, and C. Nellåker, "Turning a Blind Eye: Explicit Removal of Biases and Variation from Deep Neural Network Embeddings," in *Computer Vision – ECCV 2018 Workshops*, ser. Lecture Notes in Computer Science, L. Leal-Taixé and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 556–572.
- [11] N. G. Jablonski, "Skin color and race," *American Journal of Physical Anthropology*, pp. 1–11, 2020.
- [12] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types i through vi," *Archives of dermatology*, vol. 124, no. 6, pp. 869–871, 1988.
- [13] F. H. Rampen, B. A. Fleuren, T. M. de Boo, and W. A. Lemmens, "Unreliability of self-reported burning tendency and tanning ability," *Archives of Dermatology*, vol. 124, no. 6, pp. 885–888, Jun. 1988.
- [14] S. D. Bino and F. Bernerd, "Variations in skin colour and the biological consequences of ultraviolet radiation exposure," *British Journal of Dermatology*, vol. 169, no. s3, pp. 33–40, 2013. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bjd.12529>
- [15] A. Chardon, I. Cretois, and C. Hourseau, "Skin colour typology and suntanning pathways," *International journal of cosmetic science*, vol. 13, no. 4, pp. 191–208, 1991.
- [16] C. Cook, J. Howard, Y. Sirotin, J. Tipton, and A. Vemury, "Demographic Effects in Facial Recognition and Their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 1, Jan. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8636231>
- [17] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985, publisher: Wiley Online Library.
- [18] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. IEEE, 2002, pp. 53–58.
- [19] M. Merler, N. Ratha, R. S. Feris, and J. R. Smith, "Diversity in faces," *arXiv preprint arXiv:1901.10436*, 2019.
- [20] C. Connolly and T. Fleiss, "A study of efficiency and accuracy in the transformation from rgb to cielaab color space," *IEEE transactions on image processing*, vol. 6, no. 7, pp. 1046–1048, 1997.
- [21] M. Taylor and T. Morris, "Adaptive skin segmentation via feature-based face detection," in *Proc. SPIE 9139, Real-Time Image and Video Processing 2014, 91390P*, vol. 9139, May 2014.
- [22] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan, "A practical model for subsurface light transport," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 511–518.
- [23] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, P. E. Debevec et al., "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination," *Rendering Techniques*, vol. 2007, no. 9, p. 10, 2007.
- [24] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 145–156.
- [25] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, and P. Debevec, "Practical modeling and acquisition of layered facial reflectance," *ACM Trans. Graph.*, vol. 27, no. 5, Dec. 2008. [Online]. Available: <https://doi.org/10.1145/1409060.1409092>
- [26] T. Igarashi, K. Nishino, and S. K. Nayar, "The appearance of human skin: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 1, p. 1–95, Jan. 2007. [Online]. Available: <https://doi.org/10.1561/06000000013>
- [27] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, and H. W. Jensen, "Analysis of human faces using a measurement-based skin reflectance model," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 1013–1024, 2006, publisher: ACM New York, NY, USA.
- [28] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Feb. 2010, google-Books-ID: PTbj03bYH6kC.
- [29] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake, "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," in *ACM SIGGRAPH 2003 Papers*, 2003, pp. 770–779.
- [30] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern recognition*, vol. 41, no. 4, pp. 1350–1362, 2008, publisher: Elsevier.
- [31] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009, publisher: The Institute of Electronics, Information and Communication Engineers.
- [32] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999, publisher: Nature Publishing Group.
- [33] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *International conference on artificial neural networks*. Springer, 1997, pp. 583–588.
- [34] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1867–1874.