Improving Image Equity: Representing diverse skin tones in photographic test charts for digital camera characterization

Megan Borek; Imatest LLC; Boulder, CO, USA

Abstract

Accurate representation of diverse skin tones in photography has been a longstanding challenge due to biases toward lighter skin in traditional reference materials used for film and digital photography, such as Kodak's "Shirley" cards and the Fitzpatrick scale. These and other tools, such as the ColorChecker Classic, have offered limited ranges of skin tones and do not capture the full diversity of human skin, including variations in shades, undertones, and exposure behavior. In this study, we evaluate the application of the 10-point Monk Skin Tone Scale, developed by Harvard's Dr. Ellis Monk, to camera testing and characterization using printed skin tone charts. The Monk scale is applied to color-matched printed faces for testing cameras with facial detection capabilities. We compare the measured CIELAB values and reflectance spectra of these printed targets to those of other commonly used skin tone references, and to data measured from real human skin. Additionally, we assess the performance of these printed targets in photographed scenes in terms of exposure accuracy and color reproduction. This research identifies limitations and strengths of current printed skin tone scales and charts in representing actual human skin tones, and introduces a novel solution for improving equitable camera calibration and characterization protocols.

Background

Historically, photographic technology has been developed with a bias toward lighter skin tones, a limitation that has persisted from the era of film photography into the digital age. Early color films were chemically optimized to accurately render light skin tones, leading to significant exposure issues when capturing darker skin tones [1]. This bias was perpetuated through the widespread use of the "Shirley cards," which featured a Caucasian woman and were used as reference standards for film processing. These cards established a photographic norm that prioritized lighter skin, contributing to underexposure, poor contrast, and loss of detail in images of people with darker skin tones [1]. While the transition to digital photography has reduced some of these disparities, biases persist, with many cameras still struggling to correctly capture darker skin tones, particularly in complex lighting scenarios.

These shortcomings not only impact image quality but also perpetuate racial biases in photographic and machine learning applications. Addressing these biases is crucial for developing more equitable camera and image processing technologies that better serve diverse populations. This research seeks to provide insights into how printed skin tone targets can enhance camera testing and calibration, leading to fairer imaging practices across all skin tones.

Introduction

This research addresses two key challenges in the photographic testing of skin tones: (1) the limited diversity of skin tones represented in commercially available photographic test targets, and (2) the absence of standardized evaluation methods for automatic image processing driven by facial recognition in consumer cameras.

Fig. 1 demonstrates the impact of facial detection on automatic exposure adjustment in a consumer web camera. When the scene includes "Richard", a mannequin with a darker skin tone (Fig. 1(a)), the camera increases the exposure noticeably compared to a scene without a detectable face (Fig. 1(b)) or one featuring "Alexis" a mannequin with a lighter skin tone (Fig. 1(c)). As a result, the face in Fig. 1(a) is well exposed, but the global adjustment causes the rest of the scene to be overexposed.

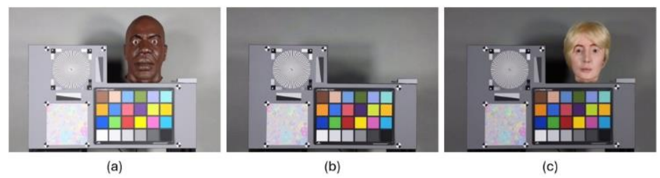

Figure 1. *Three test scenes captured with a consumer web camera under identical lighting conditions of 250 Lux, 6200 CCT. Includes (a) "Richard" mannequin with dark skin tone, (b) no mannequin, and (c) "Alexis" mannequin with light skin tone.*

The mean CIEDE2000 color difference values, calculated from the ColorChecker patches in each scene, are summarized in Table 1. As expected, the highest mean error occurs in the scene containing Richard due to facial detection-driven exposure adjustments, which overexpose the ColorChecker. Notably, the scene containing Alexis exhibits a lower mean ΔE value than the scene without a detectable face, indicating an improvement in color accuracy when a face with a lighter skin tone is placed in a scene.

Table 1: Mean CIEDE2000 of ColorChecker Patches in Mannequin Scenes

Dark Skin Tone	No Mannequin	Light Skin Tone
18.4	9.2	5.9

This experiment demonstrates the importance of practical testing in the development of cameras and image processing algorithms that adapt based on detected skin tones. To ensure accurate and equitable performance, these technologies must be evaluated using diverse, representative faces that reflect real-world scene content. However, existing camera test targets often lack both the necessary skin tone diversity and the ability to challenge facial detection algorithms. To address this gap, we present a novel set of face charts designed to represent a broader range of skin tones while effectively testing facial detection performance.

The Monk Skin Tone Scale

Developed by Ellis Monk in collaboration with Google, the Monk Skin Tone (MST) scale [2, 3] consists of 10 distinct skin tone categories, ranging from very light to very dark. Unlike previous classification systems, such as the Fitzpatrick Scale [4], which was originally designed for dermatological purposes, the MST scale was developed to address biases in computer vision and imaging technologies, ensuring more equitable representation across diverse skin tones.

As an existing scale with a wide range of tones and readily available color value information, the MST scale was an ideal candidate for translation into a printed form for this research.

Methods

Applying the Monk Skin Tone Scale to Print

A set of printed test targets featuring detectable human faces representing the 10 tones of the Monk Skin Tone (MST) scale was created (See Fig. 2). Rather than using images of real human subjects, a generative artificial intelligence (AI) tool, Generated Photos [5], was employed to efficiently produce a diverse set of synthetic faces with skin tones closely corresponding to the MST scale. To improve color accuracy and ensure a closer match to the Monk tones, additional color adjustments were applied to the generated images in Adobe Photoshop before finalizing them for the printing process [6].

A key question in this study is whether these skin tone targets can effectively replace actual human subjects in photographic testing—specifically, how well these simulated skin tones compare to real human skin in both digital reproduction and physical prints. The properties of these charts were analyzed with various methods shown in the following sections.

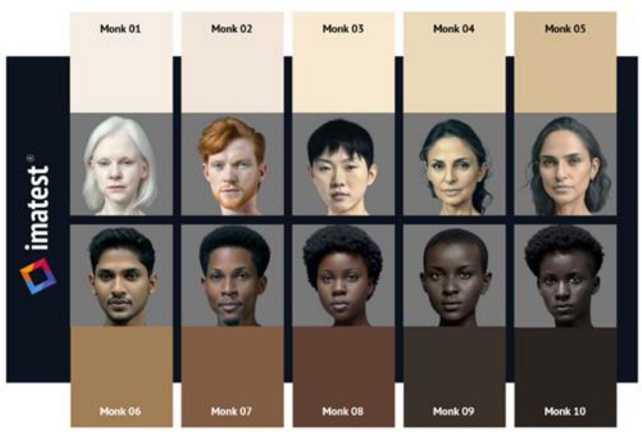

Figure 2. *Ten simulated faces with a range of skin tones based on the Monk Skin Tone Scale were created using generative AI and printed for data capture.*

Spectral Reflectance Analysis

Spectral reflectance in the visible range (400-700nm) was measured using an Avantes AvaSpec-ULS2048XL spectrometer with an AvaLight-HAL-S-Mini light source. Measurements were taken from a neutrally toned area on each face—either the forehead or cheek—avoiding regions with simulated shading or highlights. The captured reflectance spectra are shown in Fig. 3.

For comparison, Fig. 4 presents reflectance spectra from the Rochester Institute of Technology (RIT) Munsell Color Science Lab (MCSL) Lippman 2000 Dataset [7], which includes facial skin measurements from 34 subjects across 5 races [8]. Notably, the spectra of real human skin exhibit lower reflectance at shorter wavelengths compared to the simulated printed skin. Additionally, the overall higher reflectance across the visible spectrum, particularly in lighter tones, suggests that the printed targets appear brighter than natural human skin. These differences highlight potential limitations in using printed skin tone targets for color accuracy assessments in imaging applications.

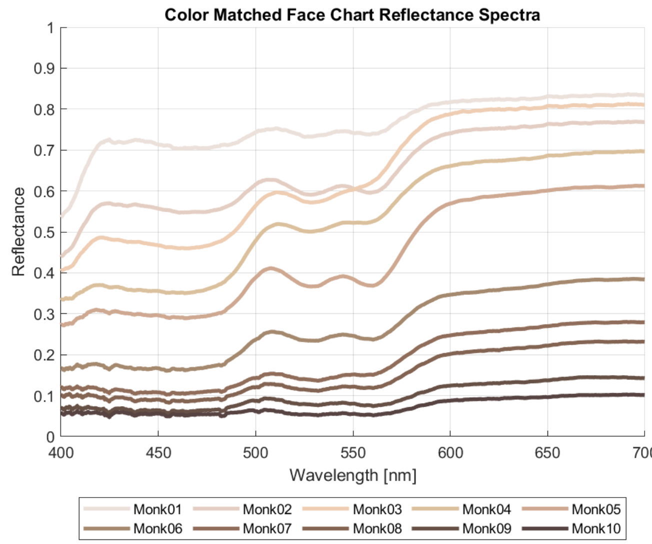

Figure 3. *Visible reflectance spectra of each color-matched face chart.*

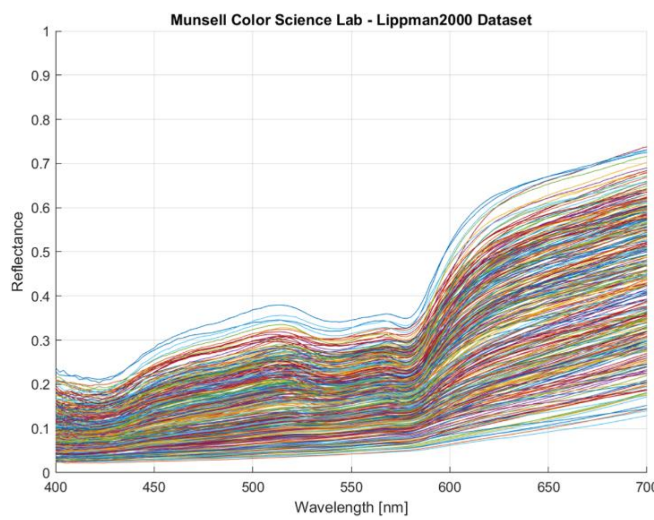

Figure 4. *Visible reflectance spectra of actual human skin, extracted from the Munsell Color Science Lab Lippman2000 Dataset.*

Waveform Analysis

Higher brightness of the Monk tones is further evidenced by the IRE (International Radio Engineers) waveforms for scenes containing the printed faces. These waveforms—originally developed for early television and still widely used in modern cinematography—visualize the brightness levels of a video signal or frame on a waveform monitor [9]. IRE values range from 0 (pure black) to 100 (pure white). A general guideline in cinematography and color grading is for skin tones in a naturally lit scene to fall between about 40 and 70 IRE for the Rec. 709 (ITU-R 709) color space, depending on the skin tone of the subject(s) in the scene [9, 10].

The IRE waveforms for a subset of the simulated faces are shown in Fig. 5, revealing brightness levels that exceed the expected range of 40-70 IRE. Brief additional experimentation demonstrated that when an actual human subject was placed in a scene alongside a printed face that they felt best matched their skin tone, the printed target appeared overexposed when the scene was properly exposed for the real human face. This discrepancy can limit the usability of these charts, which are intended to be a more accessible alternative to human subjects, if they do not accurately represent skin tones and their exposure behavior.

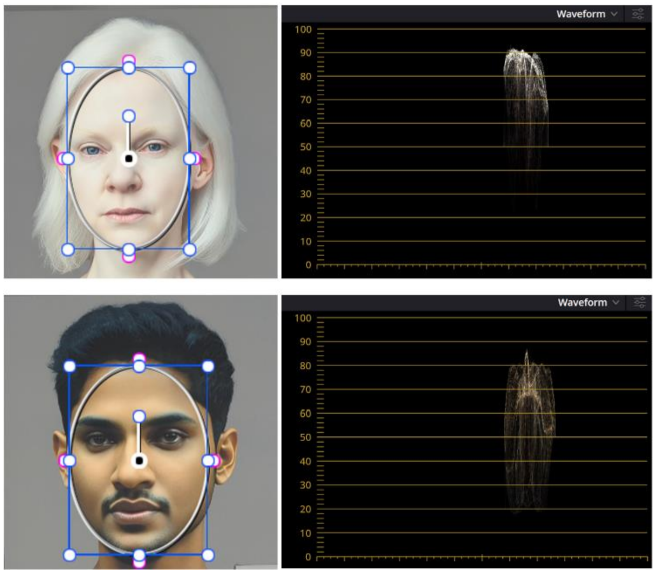

Figure 5. *IRE waveforms (right) for ROIs (left) of Monk 01 (top) and Monk 06 (bottom) color-matched face charts. Scenes illuminated at 250 Lux, 6200 CCT, captured with Google Pixel 6 Pro.*

Vectorscope Analysis

We also analyze the color characteristics of the generated faces using vectorscopes, a tool frequently used in video post-production for visualizing color information in an image. Vectorscopes represent hue as an angular position (where 0° is on the positive x-axis) and saturation as the radial distance from the center, typically in a polar YUV space.

One commonly referenced feature of these scopes in color grading is the positive I-line (in-phase component) at 123°, which aligns with the orange hues to which human vision is particularly sensitive. While the use of the I-line as a strict reference for skin tone correction is debated, it remains a useful guideline, as human skin tones—regardless of complexion—generally fall near this axis under neutral lighting conditions [9, 11].

To further evaluate the MST faces, we analyze the color distribution of a circular region of interest (ROI) on each face using vectorscope readings in DaVinci Resolve, as shown in Fig. 7. The results, captured under neutral lighting with a Google Pixel 6 Pro, indicate that the simulated skin tones—particularly those in the light to mid-tone range—exhibit noticeable shifts toward green, extending well beyond the I-line.

Improving Printed Skin Tones

The previous tests reveal that the Monk tones do not accurately represent the color and spectral properties of human skin in printed form. Two primary issues arise: (1) the hues lean too yellow/green, and (2) their brightness is significantly higher than that of actual human skin. In terms of the CIELAB color space, this corresponds to the L* (lightness) component being too high and the a* (green-red) component being too low compared to real skin tones.

To correct these discrepancies, the image files were further adjusted in Adobe Photoshop. Working in CIELAB space, the L* component is lowered, and the a* component is increased for each face. Adjustments were guided by comparing the modified L*a*b* values to L*a*b* values derived from the MCSL spectral curves (See Section *L*a*b* Comparison* for details on the conversion to CIELAB). This process was iterative, involving several rounds of digital modifications, printing, and measurement of the prints to improve alignment with real skin tones.

Spectral Reflectance Analysis

New spectral reflectance measurements for the color-adjusted prints were taken and are shown in Fig. 6. Compared to the spectra of actual human skin (see Fig. 4), the adjusted tones now exhibit similar spectral shapes and reflectance values across the visible spectrum, suggesting an improved representation of real skin tones. However, it is important to note that these prints remain susceptible to metamerism, meaning their appearance may change under different lighting conditions.

A key characteristic of human skin reflectance is the presence of two central valleys at about 540 nm and 577 nm, which correspond to the absorption bands of hemoglobin. These influence the perception of skin tones under natural and artificial lighting. While the printed charts reproduce a similar spectral shape, these valleys are shifted to around 530 nm and 560 nm, respectively. This shift highlights the challenges of accurately replicating hemoglobin absorption using traditional CMYK printing methods, as the limited pigments cannot perfectly mimic the selective absorption properties of human skin.

This limitation is particularly relevant for applications requiring precise skin tone reproduction. Future improvements may involve printing on, or layering, different substrates.

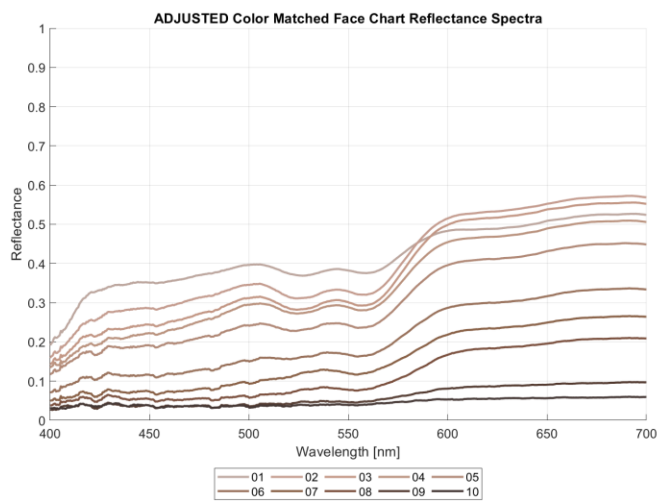

Figure 6. *Visible reflectance spectra of each adjusted face chart.*

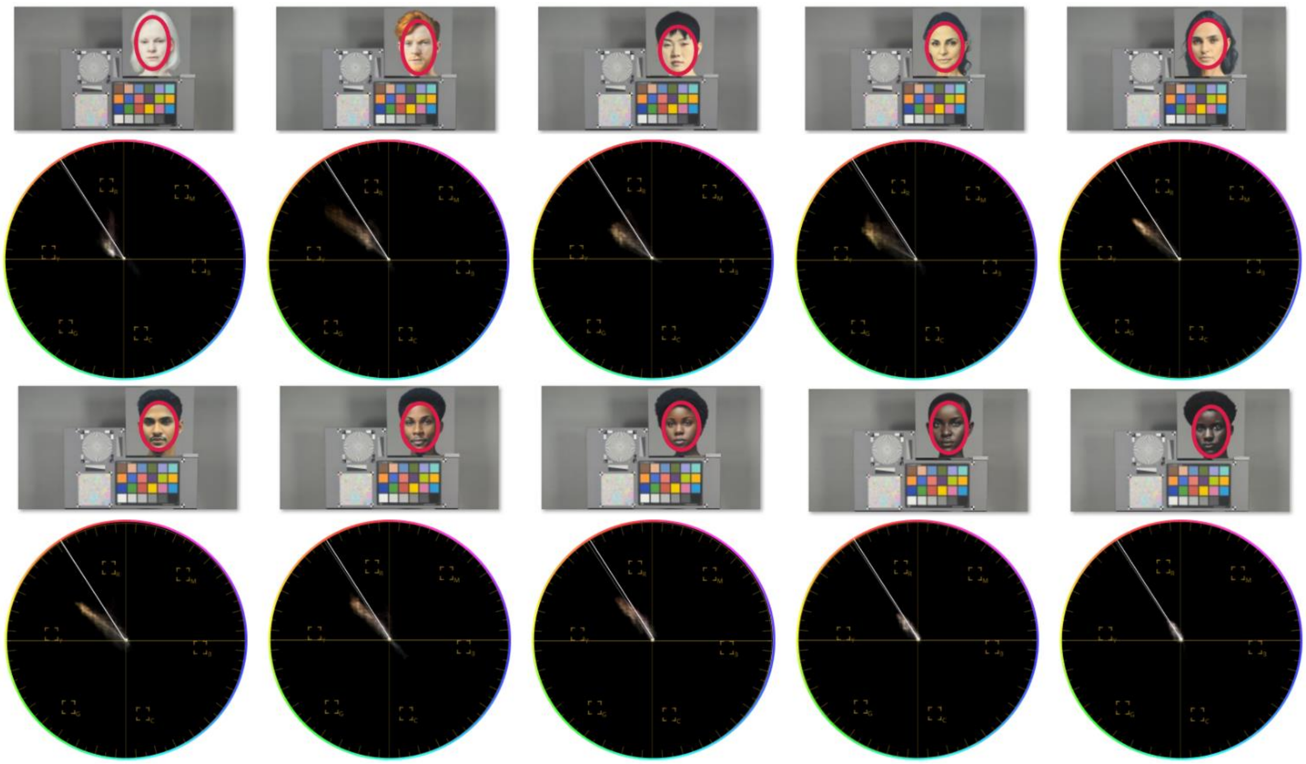

Figure 7. *DaVinci Resolve vectorscopes of the Monk tone color-matched face charts. Scenes illuminated at 250 Lux, 6200 CCT, captured with Google Pixel 6 Pro.*

Waveform Analysis

Initial test images of the adjusted prints were captured with a Google Pixel 9 Pro for analysis. Viewing the IRE waveforms for a subset of the faces, shown in Fig. 8, reveals a significant decrease in brightness for the adjusted faces, compared to the original faces color-matched to the Monk tones (see Fig. 5). Additional test captures remain to be taken in order to properly compare the charts in identical lighting and capture conditions.

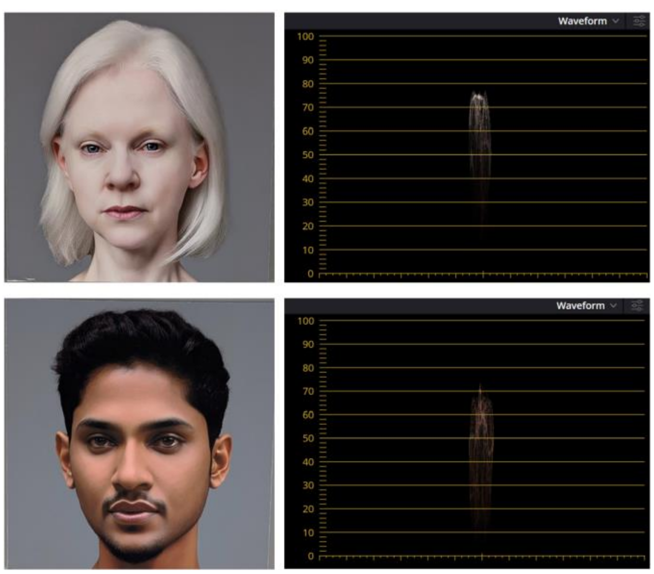

Figure 8. *IRE waveforms (right) for ROIs (left) of adjusted Monk 01 (top) and Monk 06 (bottom) face charts.*

Vectorscope Analysis

Vectorscope analysis of the adjusted targets (Fig. 9) shows a notable improvement in hue accuracy compared to the original Monk tone targets (Fig. 7). ROIs in the facial areas now appear on or slightly above the I-line, aligning more closely with how actual human skin typically registers on a vectorscope.

However, achieving a more comprehensive representation of skin diversity requires further refinement. While the current adjustments bring the tones closer to real-world skin hues, they do not yet fully account for the natural range of undertones present in human skin. To address this, additional modifications are planned to expand the range of undertones represented. The goal is to ensure that the adjusted targets encompass a broader spectrum to better reflect the natural variation in skin tones, from cooler (blue/olive) to warmer (red/golden) undertones.

*L*a*b* Comparison*

Multiple new spectral reflectance measurements were taken across each adjusted face chart and converted to L*a*b* values (Fig 10). Spectral reflectance spectra were converted to XYZ using CIE 1931 color matching functions for a 2° observer, and then to CIELAB for the D50 standard illuminant. The original MST L*a*b* values are plotted (Fig. 11), along with L*a*b* values derived from the MCSL spectral dataset (Fig. 12) and the Pantone SkinTone Guide (Fig. 13) which contains 138 shades based on real skin measurements, and covers a broad range of lightness levels and undertones [12].

Measurements from the adjusted charts show that the new range of tones more closely aligns with the gamut of real skin color. Compared to the original MST values, the adjustments result in a

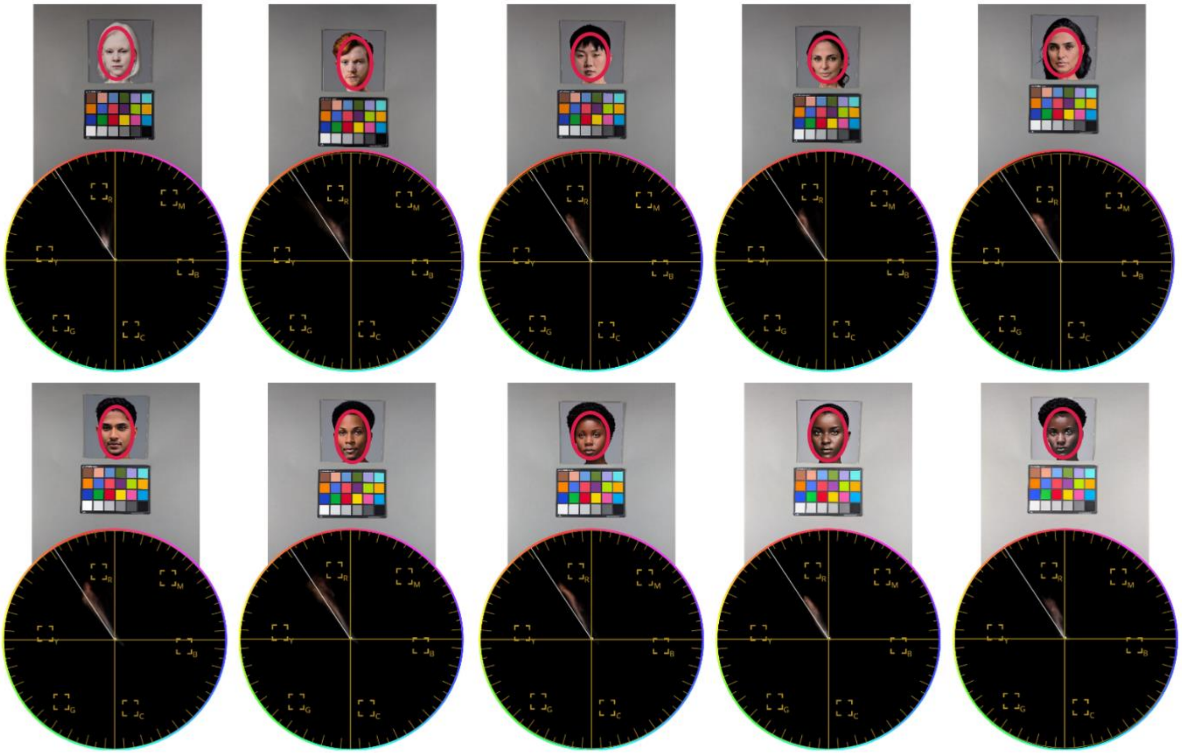

Figure 9. *DaVinci Resolve vectorscopes of the color-adjusted face charts*

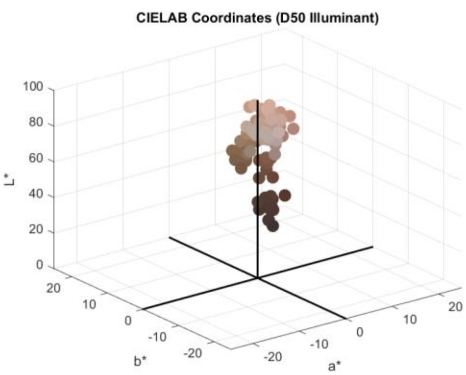

Figure 10. *L*a*b* (D50, 2°) values of adjusted skin tones.*

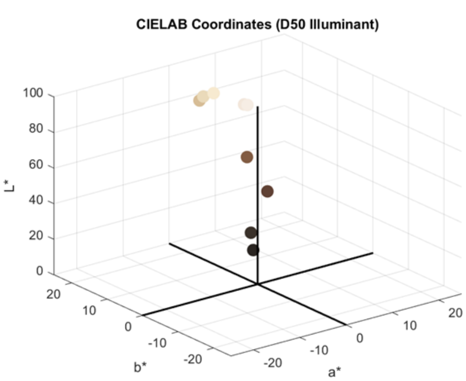

Figure 11. *L*a*b* (D50, 2°) values of original Monk Skin Tones.*

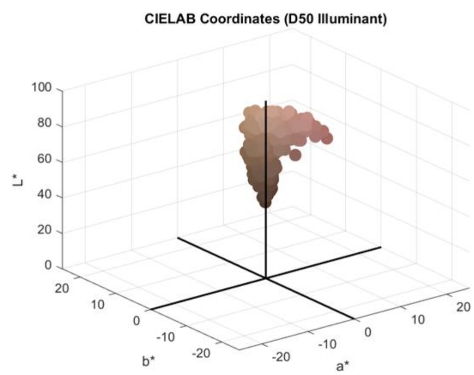

Figure 12. *L*a*b* (D50, 2°) values derived from MSCL reflectance spectra*

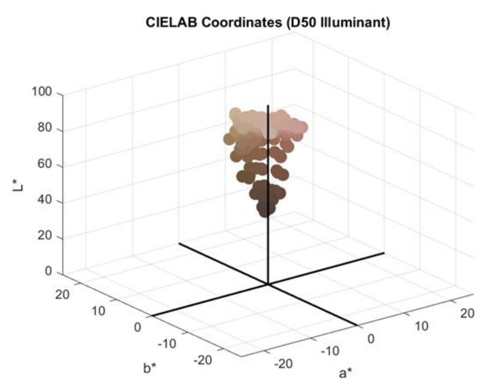

Figure 13. *L*a*b* (D50, 2°) values from Pantone SkinTone Guide*

more accurate distribution of tones that better reflect natural variations in skin color.

Conclusion

This study highlights the limitations of existing printed skin tone targets, particularly those based on the Monk Skin Tone (MST) scale, in accurately representing the color and spectral properties of human skin. Initial evaluations revealed that the MST-based printed faces exhibited hue shifts toward yellow-green and higher overall brightness compared to real skin, leading to inaccuracies in photographic testing. Spectral reflectance analysis confirmed these discrepancies, showing significant deviations from real skin reflectance, particularly in the absorption features of hemoglobin. Additionally, IRE waveform and vectorscope analyses demonstrated that the printed charts did not accurately reproduce the exposure behavior and color characteristics of human skin in imaging applications.

To address these issues, adjustments were made to the printed targets using iterative modifications in the CIELAB color space, resulting in improved alignment with real skin tones. The refined prints demonstrated closer spectral similarity to measured human skin data, though metamerism remained a challenge. Further refinements, particularly in capturing a broader range of undertones and fine-tuning spectral absorption features, are necessary for continued improvement.

The findings of this research emphasize the importance of developing more accurate and diverse skin tone references for imaging applications, particularly in camera calibration, color grading, and machine learning for facial recognition. By improving printed skin tone targets, we can contribute to fairer and more equitable imaging practices, reducing biases in consumer cameras and advancing the representation of diverse skin tones in digital imaging. Future work will explore enhanced printing techniques, additional adjustments to spectral properties, and expanded datasets to ensure even more accurate representation of human skin across a variety of imaging scenarios.

References

[1] M. Zhang, "How Color Film was Originally Biased Toward White People," PetaPixel, 19 September 2015. [Online]. Available: https://petapixel.com/2015/09/19/heres-a-look-at-how-color-film-was-originally-biased-toward-white-people/. [Accessed September 2024].

[2] E. Monk, "The Monk Skin Tone Scale," 4 May 2023. [Online]. Available: https://osf.io/preprints/socarxiv/pdf4c_v1. [Accessed September 2024].

[3] Google Research, "Developing the Monk Skin Tone Scale," Google Research, [Online]. Available: https://skintone.google/the-scale. [Accessed August 2024].

[4] W. H. Ward, F. Lambreton and N. Goel, "Clinical Presentation and Staging of Melanoma," in *Cutaneous Melanoma: Etiology and Therapy*, Brisbane (AU): Codon Publications, 2017.

[5] "AI Human Generator," Generated Media, Inc., [Online]. Available: https://generated.photos/.

[6] M. Borek, "Evaluating Camera Performance in Face-Present Scenes With Diverse Skin Tones," *Electronic Imaging,* vol. 36, pp. 260-1 - 260-5, 2024.

[7] "Lippman2000 Spectral Imaging Database," RIT Munsell Color Science Lab, 3 October 2013. [Online]. Available: https://www.rit-mcsl.org/Lippman2000/Spectral/SpectralData/HumanFace/Spectral_Data/.

[8] Q. Sun and M. D. Fairchild, "Statistical Characterization of Face Spectral Reflectances and Its Application to Human Portraiture Spectral Estimation," *Journal of Imaging Science and Technology,* vol. 46, pp. 498 - 506, 2022.

[9] B. Brown, in *Cinematography Theory & Practice, 3rd Edition*, Routledge, 2016, pp. 156-161, 183-185.

[10] S. Sudhakaran, "A Quick look at Understanding IRE Values," wolfcrow, 18 December 2024. [Online]. Available: https://wolfcrow.com/a-quick-look-at-understanding-ire-values/.

[11] K. Jack, "NTSC and PAL Digital Encoding and Decoding," in *Video Demystified: A Handbook for the Digital Engineer, 5th Edition*, Newnes, 2007, p. 432.

[12] "Pantone SkinTone Guide," Pantone, [Online]. Available: https://www.pantone.com/skintone.

Author Biography

Megan Borek received her BS in Imaging Science from the Rochester Institute of Technology (2022), where her studies focused on motion picture science, remote sensing, and computer vision. She is currently an Imaging Scientist at Imatest LLC in Boulder, CO, where she focuses on imaging standards, software development, and test lab services. Her work involves evaluating camera performance, developing testing methodologies, and advancing image quality assessment techniques for a range of applications.