Received: 8 December 2017 | Revised: 31 December 2018 | Accepted: 2 January 2019

DOI: 10.1002/col.22365

WILEY

RESEARCH ARTICLE

Assessing color performance of whole-slide imaging scanners for digital pathology

Wei-Chung Cheng[1] | Firdous Saleheen[1,2] | Aldo Badano[1]

[1]US Food and Drug Administration, Center for Devices and Radiological Health/Office of Science and Engineering Laboratories/Division of Imaging, Diagnostics, and Software Reliability, Silver Spring, Maryland

[2]Engineering Department of EDDA Technology, Inc., Princeton, New Jersey

Correspondence
Wei-Chung Cheng, US Food and Drug Administration, CDRH/OSEL/DIDSR, Silver Spring, MD 20993.
Email: wei-chung.cheng@fda.hhs.gov

Abstract

The color performance of two commercial whole-slide imaging (WSI) scanners was compared against the ground truth and a hypothetical monochrome scanner. Three biological tissue slides were used to test the WSI scanners. A multispectral imaging system was developed to obtain the color truth of the biological tissue slides at the pixel level. The hypothetical monochrome scanner was derived from the color truth as a lower bound for comparison. The CIEDE2000 formula was used to measure color errors. Results show that color errors generated by the modern commercial WSI scanner, the legacy commercial WSI scanner, and the monochrome WSI scanner are in the range of [8.4, 13.0], [18.0, 26.33], and [17.4, 17.6] ΔE_{00}, respectively. The legacy commercial WSI scanner was outperformed by not only the modern commercial WSI scanner but also by the hypothetical monochrome scanner.

KEYWORDS

color measurement, color reproduction, digital pathology, multispectral imaging, whole-slide imaging

1 | INTRODUCTION

Thanks to technological advances in digital microscopy, image compression algorithms, display quality, computer performance, and Internet connectivity, the *whole-slide imaging* (WSI) scanner is emerging as a replacement for the light microscope that has been used for decades by pathologists to review tissue slides. Using a light microscope, the pathologist examines a tiny region of the glass slide that is magnified by the objectives and eye pieces. Since the region is extremely small, the pathologist needs to maneuver the stage that carries the glass slide translationally to move from one region to another. Because the depth of field of the high-magnification objectives is shallow and the tissue sample is not always flat, continuous refocusing is required to retain a clear view. The field of view and the magnification can be changed by switching to different objectives. The pathologist views the tissue sample illuminated by the light source directly. Since there is no intermediate media involved, the view is updated in real time. The human visual system adapts to the lighting conditions inside the microscope light path.

The WSI technology provides a new paradigm of scanning, viewing, analyzing, managing, and sharing pathological images digitally. Unlike a light microscope, the WSI system pre-scans the glass slide offline. The small regions of the slide, usually hundreds of thousands, are focused and imaged one by one. Then the small regions are stitched together to form a large whole-slide image, which consists of millions of pixels. The large amount of pixel data requires effective compression algorithms to reduce the size so that the data can be processed, stored, and transmitted with available computer resources. The whole-slide image is then transmitted and displayed in a review workstation that might be in a remote location. The WSI display provides a larger and more ergonomic view than that of the light microscope. The pathologist interacts with the whole-slide image via

mice, keyboards, or other human interface devices rather than the stage and focus knobs. Since the image is reproduced by the WSI system, the color can look very different compared with what the light microscope shows. The response time of image update depends on the computer environment and the complexity of the whole-slide images.

Color performance was identified as an important and emergent topic by the medical imaging community.[1] As a class II medical imaging device, the WSI device marketed in the United States is regulated by the US Food and Drug Administration (FDA), the Center for Devices and Radiological Health.[2,3] Guidance for assessing the technical performance of WSI devices was released in 2016, in which color performance is one of the required system-level tests.[4] According to the guidance, the regulatory clearance of a WSI device requires the manufacturer to provide color performance test data that include the color rendering intents, test method, test result, and analysis. The test method is not defined in the guidance but left for the manufacturer to design. The first WSI device was granted approval by the FDA in April 2017 via the de novo premarket review pathway.

A WSI system is essentially a color reproduction system that consists of the image acquisition subsystem and the image display subsystem. In this study, only the image acquisition subsystem, that is, the WSI scanner, was evaluated. The input to the WSI scanner is a glass slide with tissue samples on it. The output from the WSI scanner is a whole-slide image. The goal of this study is to measure the color differences between the input tissue sample on the glass slide and the output whole-slide image. Based on the measurement results, the color performance of different WSI scanners can then be compared quantitatively.

Compared with the other color reproduction systems, the challenge of assessing WSI scanners is that the scene, a field of view of the glass slide, is in the microscopic range that needs to be magnified by 100 to 600 times to be visible to the human eye. The resolution of commercial WSI scanners ranges from 110 nm/pixel (Zeiss Axio Scan.Z1, 40×) to 250 nm/pixel (Aperio ScanScope CS, 40×). In contrast, for a spectroradiometer with a high-magnification lens through the smallest aperture (Photo Research PR730, SL-1X lens, 0.01° aperture), the measurement spot size is 89 μm, about 800 times larger than that of WSI scanners. Thus, a conventional spectroradiometer is inadequate for determining the color truth for assessing WSI scanners unless the target is uniform and larger than the measuring spot.

Nevertheless, a glass slide with known or measurable color truth is still required for evaluating or calibrating WSI scanners. In the literature, researchers have used different types of man-made color targets to represent biological tissue slides. These man-made color targets consist of a handful of color patches of a size that is both measurable by spectrophotometers and scannable by WSI scanners. The commercial color patches are made of materials with consistent optical characteristics so that the color targets can be manufactured and distributed in volume. The materials used for color targets include photographic film,[5,6] optical filters,[7,8] and tissue-like substrates.[9]

Visible large-sized color patches were mounted on a glass slide as the color target to test the WSI device.[8] The output WSI image and the color target were visually compared side-by-side in front of the display to determine the color reproducibility. This method provides a quick and intuitive way of detecting a malfunctioning WSI device for quality control purposes. However, the decision might be subjective and cannot serve as a quantitative measure for comparing devices.

Photographic film-based color targets have also been used as color targets.[5,6] More than 100 measurable and customizable color patches constitute a wide range of color gamut for testing the WSI device's color response. However, the spectral transmittance of a photographic film is known to differ from that of biological tissues and might lead to inaccurate assessment.

Recent work describes the selection and design of a set of miniature narrow-band optical filters installed on a slide for calibrating WSI scanners.[7] Unlike the abovementioned approaches, these optical filters were highly homogeneous, spatially uniform, narrowband, consistent, and repeatable. However, these uniform color filters lacking texture increased difficulty in focusing for WSI scanners and compromised their scanning ability.

A spectrally matching color target was developed with a biopolymer material that could be stained by common histological protocols.[9] Although the color target spectrally emulated biological tissues very well, due to its laborious manufacturing method, the number of color patches on a glass slide was limited.

Although man-made color targets have the benefits of measurability and reproducibility that biological tissue samples lack, they differ from biological tissue samples in three characteristics: spectral characteristics, color gamut, and microscopic structure. First, due to the limited availability of transparent materials, it is difficult to make a color target that spectrally matches those of a biological tissue sample. The difficulty is further elevated when taking the variability in tissue preparation and staining processes into account. Second, the number of color patches mounted on a color target is limited and far less than those presented in a biological tissue slide. Finally, the color patches, which must be reasonably large to be measurable and reproducible, do not mimic the microscopic structure of biological tissue slides. As a result, the color performance degraded by the optics cannot be assessed. For example, color errors caused by chromatic aberration will not be observed inside large color patches.

To overcome the two hurdles that man-made color targets were handicapped, spectral matching and microscopic structure, our approach is to use real biological tissue slides as the

targets and measure the color truth embedded within the microscopic structures using multispectral imaging techniques.[10]

In this article, multispectral imaging refers to the technique that measures the spectral power distribution of each pixel in the image so that, specifically in this work, the colorimetric measure (ie, CIEXYZ tristimulus values) of each pixel can be determined. The measured spectral power distribution must have a sufficient spectral resolution, that is, the number of bands in the visible wavelength range, to calculate the tristimulus values accurately. For example, an imaging device that uses only three narrow bands to generate color images might also be referred to as multispectral without the capability of faithfully reproducing original colors.

In color imaging, multispectral imaging is the standard means for obtaining the ground truth of the original scene. The key component of multispectral imaging is how to decompose the broadband visible light into a set of narrowbands, which can be implemented in different ways.[11] One popular method is using narrowband optical filters to select the desired wavelength from the broadband white light. A set of dichromatic color filters with preselected wavelengths can be installed on a motorized wheel for automated imaging. Alternatively, a tunable color filter such as a liquid crystal tunable filter (LCTF) can be controlled dynamically to select a specific wavelength. Although LCTF is a more efficient and flexible apparatus compared with motorized color filters, the spectral responses are not uniform due to the optical properties of liquid crystals. The transmittance of a shorter wavelength (eg, blue light) might be too low and might create noise problems for the image detector. Both fixed and tunable filter methods share the potential problem of camera noises originated from insufficient light. The narrowband filters reduce the illumination greatly, so a regular light source designed for a direct view by the human eye is inadequate. In this study, a microelectromechanical system's (MEMS) tunable light source was used to replace the tungsten lamp of a light microscope. The tunable light source provides the capability of finely tunable bandwidth, synthesizable spectrum, and fast response.

The measurement results from the multispectral imaging system were used to generate not only the color truth but also the monochrome image. The monochrome image was derived from the color truth by stripping the chromaticity information. The concept of employing a monochrome image can be considered as "color falsity," which counters the concept of color truth. The purpose of using the dual nature of color truth and color falsity was to create an interval of quantitative color differences on which the color performance of WSI scanners can be scaled.

In this study, only the image acquisition subsystems of the WSI devices were assessed. The display subsystems, whose color performance can be affected by too many factors, were evaluated separately in another study.

2 | MATERIALS AND METHODS

The workflow of the methodology is illustrated in Figure 1. A biological tissue sample mounted on a glass slide was imaged by both the multispectral imaging system and the WSI scanner under test. The output images from both systems were registered and converted into the CIELAB (CIE L*a*b*) color space to calculate the ΔE_{00} color difference for each pixel. The pipelines for the multispectral imaging and the WSI scanners are described in the following sections.

3 | COLOR TRUTHING WITH MULTISPECTRAL IMAGING

3.1 | Device description

The multispectral imaging system was implemented by retrofitting a conventional light microscope. It comprised four components—microscope, tunable light source, camera sensor, and control software. The hub of the multispectral image system was an upright light microscope (AxioPhot 2, Carl Zeiss Microscopy, NY) in bright-field mode with a 10× objective (Carl Zeiss A-Plan 10×/0.25 Ph1). A motorized

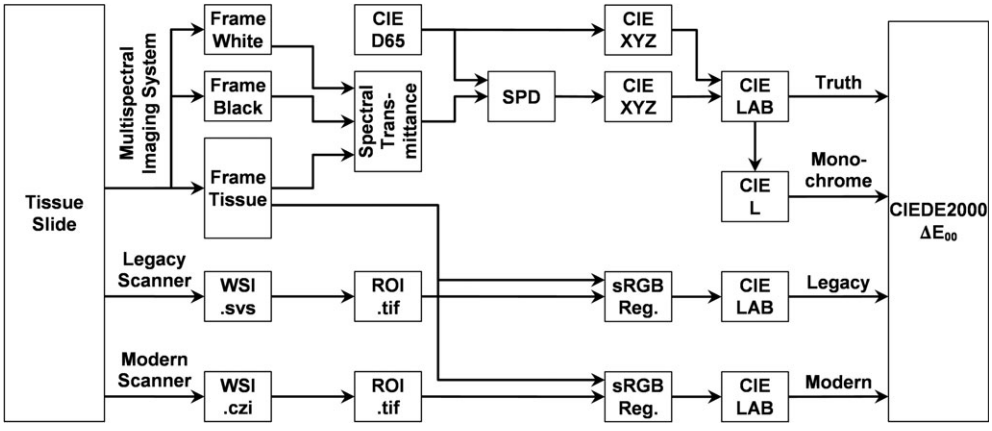

FIGURE 1 Methodology overview

XY-stage (MAC 6000, Ludl Electronic Products, Hawthorne, NY) translationally moved the glass slide under the objective to select the desired region of interest (ROI). The glass slide was illuminated by, in lieu of a conventional tungsten halogen lamp, a tunable light source (OL490 Agile Light Source, Gooch and Housego, TX). On the detector side, a scientific monochrome charge-coupled device (CCD) camera (Grasshopper3 9.1 MP Mono USB3 Vision, Point Grey Research, BC, Canada) calibrated to a linear response was used to measure the luminance of each pixel in the field of view. After mounting the light source and camera, Kohler illumination was attained by refocusing the condenser (Zeiss achromatic-aplanatic condenser system, aperture 0.9) accordingly. The image was focused manually with 550 nm (green) light.

Inside the tunable light source, the broadband white light from a xenon lamp was dispersed into various wavelengths by prisms. In this study, the 150-μm aperture was selected to generate the narrowest bandwidth. A fast-switching MEMS-based Digital Light Processor (Texas Instruments, Dallas, TX) with 1024 columns was software-controlled to reflect a set of selected wavelengths. The mapping between the 1024 columns and the wavelength was nonlinear and needed to be determined at factory as a calibration file. The factory software looked up the calibration file and actuated the corresponding columns based on the user's choice of wavelength. The wavelengths reflected by the actuated columns were then combined and delivered through a liquid light guide. The liquid light guide was coupled with the light microscope with a collimating adapter (LLG5A4-A, Thorlabs, Newton, NJ).

The pixel count of the camera was 3376 × 2704 at 9 fps. The size of the CCD sensor (ICX814, Sony Electronics, Park Ridge, NJ) was 1 in., which covers a major portion of the field of view of the microscope. The resolution of the microscope system when using the 10× objective was 370 nm per pixel. The tunable light source, the motorized stage, and the camera were all controlled by programs written in Matlab 2015b (MathWorks, Natick, MA) running in the Microsoft Windows 7 Professional 64-bit environment.

3.2 | Principles

The goal of the multispectral imaging system was to measure the spectral transmittance, T, of the glass slide for each pixel. The glass slide was illuminated by the light source with spectral power distribution L. The spectral sensitivity of the camera, including both the sensor and optics, is C, which was also dependent on the camera parameters such as the exposure, gain, and brightness settings. The camera detects the luminance of the target Y and reports in the digital count D. The detected luminance Y is proportional to C, L, and T over the visible light wavelength λ:

$$Y \propto \int C(\lambda) \cdot L(\lambda) \cdot T(\lambda) \tag{1}$$

By fixing the settings of the camera and the spectrum of the light source, C and L can be considered constant so the detected luminance at a specific wavelength is proportional to the transmittance of the tissue sample on the glass slide only:

$$Y(\lambda) \propto T(\lambda) \tag{2}$$

If the camera has a linear response between the detected luminance Y and the output digital count D, then the relationship between the transmittance and the digital count can be expressed as a linear function:

$$D(\lambda) = k_1 \cdot T(\lambda) + k_0 \tag{3}$$

The two coefficients k_0 and k_1 can be found by measuring two different glass slides with known transmittance. In this study, a transparent target ($T = 1$) and an opaque target ($T = 0$) were used.

$$D_{\mathrm{Max}}(\lambda) = k_1 \cdot 1 + k_0 = k_1 + k_0 \tag{4}$$

$$D_{\mathrm{Min}}(\lambda) = k_1 \cdot 0 + k_0 = k_0 \tag{5}$$

Then the transmittance can be calculated via interpolation:

$$T(\lambda) = \frac{D(\lambda) - D_{\mathrm{Min}}(\lambda)}{D_{\mathrm{Max}}(\lambda) - D_{\mathrm{Min}}(\lambda)} \tag{6}$$

Recall that the spectral power distribution of the tissue sample, S, is the product of the spectral power distributions of the light source and the transmittance. Although the resulting spectral transmittance is the ground truth of the glass slide, it cannot serve as the color truth for comparison until a light source is applied. In this study, the International Commission on Illumination (Commission internationale de l'éclairage, CIE) D65 illuminant was used.

$$S(\lambda) = L_{\mathrm{D65}}(\lambda) \cdot T(\lambda) \tag{7}$$

The CIEXYZ tristimulus XYZ values were calculated by

$$X = \int S(\lambda)\bar{x}(\lambda)d\lambda \tag{8}$$

$$Y = \int S(\lambda)\bar{y}(\lambda)d\lambda \tag{9}$$

$$Z = \int S(\lambda)\bar{z}(\lambda)d\lambda, \tag{10}$$

where $\bar{x}$, $\bar{y}$, and $\bar{z}$ are the CIE 1931 color matching functions.

Then the tristimulus values of the CIE D65 illuminant were calculated similarly and used as the reference white, $X_nY_nZ_n$, to calculate the CIELAB $L^*a^*b^*$ values.

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \tag{11}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{12}$$

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{13}$$

$$f(t) = \begin{cases} \sqrt[3]{t} & \text{if } t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29} & \text{otherwise} \end{cases} \tag{14}$$

$$\delta = \frac{6}{29} \tag{15}$$

The CIELAB $L^*a^*b^*$ values would be used as the truth to evaluate the WSI scanners. For visualization and presentation purposes, the CIELAB data were also converted into the standard red green blue (sRGB) color space to reconstruct the truth image.[12] Notice that the sRGB color space was not used to process the truth data because some eosin-stained shades with high lightness values are out of the sRGB color gamut. Nevertheless, many commercial WSI scanners use sRGB to store and display images. In this study, the CIE D65 illuminant was used to establish the truth. Alternative illuminants can be used for specific viewing environments or applications.

3.3 | Workflow for multispectral imaging

The conceptual procedure of multispectral imaging is described in the following pseudocode.

```
Adjust the camera settings
 For each wavelength λ=380:10:780
    Adjust the tunable light source for λ
    Image the opaque target to obtain D_min for T=0
    Image the transparent target to obtain D_max
    for T=1
    Measure the target slide to obtain D for each
    pixel
    Linearly interpolate (D,T) from (D_min,0) and
    (D_max,1)
 End
 Reconstruct spectral transmittance for each
pixel
```

The spectra of the light source are shown in Figure 2. The set of 41 images were compared with both flat-field bright and dark images to calculate the spectral transmittance. The flat-field bright image was acquired by imaging a blank area on a glass slide with a coverslip mounted. The camera exposure settings were adjusted such that the bright pixel values did not saturate. These pixel values, D_{max}, which may vary among pixels due to spatial nonuniformity, indicate the 100% transmittance, which already included the effects of the glass slide and the coverslip. The flat-field dark image was acquired by imaging a black plate that does not transmit light. These pixel values, D_{min}, indicate the 0% transmittance. For each wavelength, each pixel reading was linearly interpolated between the bright and dark pixel values to obtain the per-wavelength transmittance (Equation 6). After collecting the per-wavelength transmittance data from 41 images, the complete spectral transmittance of each pixel was reconstructed (Figure 1, spectral transmittance). The spectral transmittance was multiplied by the spectrum of the CIE D65 illuminant (Equation 7) to obtain the spectral power distribution (Figure 1, SPD). The spectral power distributions of the tissue and the reference white were converted to CIEXYZ (Equations 8-10), which generated CIELAB (Equations 11-15). The procedure was implemented in Matlab conceptually but not verbatim because data vectorization was required to minimize the computation time for large images.

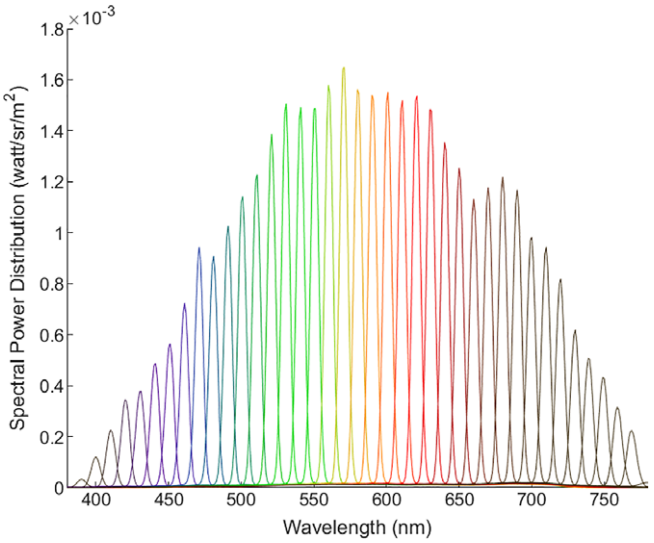

FIGURE 2 The spectral power distributions of the 41 bands used to obtain the multispectral images shown in the corresponding hue

3.4 | WSI scanners

The workflow for the WSI scanners is illustrated along the lower arms in Figure 1. The glass slide was scanned to generate a WSI file in the proprietary file format, which contains the whole-slide image at different magnification levels, label image, thumbnail image, and metadata. After scanning, the ROI, which was determined based on the field of view acquired by the multispectral imaging system, was retrieved, registered, and exported to the tagged image file format (TIFF). The red, green, and blue (RGB) pixel data were then retrieved from the TIFF file and converted from sRGB to CIELAB color space for comparison with the truth.

The three WSI scanners, namely *legacy*, *modern*, and *monochrome*, evaluated in this study are described as follows.

3.4.1 | Legacy WSI scanner

The legacy WSI scanner is part of a commercial WSI system, Aperio ScanScope CS (Leica Biosystems, Inc., Buffalo Grove, IL 60089, USA). The scanner generated WSI images in the proprietary SVS file format (Figure 1, "WSI.svs"). The ROI was retrieved and exported with the factory review software (Aperio ImageScope 12.2.2.5015) with the default settings (Figure 1, "ROI.tif"). sRGB was the default color space for the factory display subsystem. The color rendering intent was unknown.

3.4.2 | Modern WSI scanner

The modern WSI scanner is part of another commercial WSI system, Zeiss Axio Scan.Z1 (Carl Zeiss Microscopy, NY). The scanner generated WSI images in the proprietary CZI file format. The ROI was retrieved and exported using the factory review software (Zeiss ZEN2 ver. 2) with the default settings (Figure 1, "WSI.czi"). sRGB was the default color space for the factory display subsystem (Figure 1, "ROI.tif"). The color rendering intent was unknown. The modern WSI scanner was manufactured 6 years after the legacy WSI scanner.

3.4.3 | Monochrome WSI scanner

The monochrome WSI scanner is a hypothetical device that was derived from the multispectral imaging system. In the multispectral imaging workflow, after calculating the CIELAB $L^*a^*b^*$ values, the a^* and b^* values were set to zero to generate a new image that preserves only the lightness without the chromaticity (Figure 1, CIE L). The monochrome image was then compared with the truth image in the CIELAB color space to calculate the CIEDE2000 values. Notice that, as depicted in Figure 1, the output of the monochrome WSI scanner did not go through the sRGB conversion. The original intention of introducing this hypothetical device was to set a lower bound so that the performance of a commercial WSI scanner could be graded between the truth and the monochrome device.

3.5 | Image registration

The Matlab *Registration Estimator* app with the *Maximally Stable Extremal Regions* (MSER) algorithm was used to register the scanned sRGB image with the truth. One of the multispectral frames was used as the registration reference. The app generated a two-dimensional affine geometric transformation object, which was used to transform the scanned sRGB image with the *imwarp* function (Figure 1, "sRGB Reg."). The registered sRGB image was then converted into the CIELAB color space based on the sRGB standard.[12]

An interactive program was written to analyze the registration accuracy of the app. In the snapshot shown in Figure 3, the correlation coefficients on two cross sections were 0.9755 and 0.9697, which indicate adequate registration accuracy.

3.6 | CIEDE2000 comparison

For each pixel, the color difference between the WSI scanner and the truth was calculated with the CIEDE2000 formulas based on their CIELAB values (Figure 1, "CIEDE2000 ΔE_{00}"). The lengthy calculation of CIEDE2000 is omitted here but can be found in the literature.[13] Basically, the CIEDE2000 is an improved color difference model of the CIE 1976 color difference formula (CIE76), which simply calculates the Euclidean distance between two color coordinates in the CIELAB color space. The CIE 1976 color difference is not perceptually accurate because the CIELAB color space is not perfectly perceptually uniform. The nonuniformity of CIELAB was corrected by adding five correction parameters in the CIEDE2000 model.

3.7 | Histological samples

Three hematoxylin-and-eosin (H&E) stained tissue samples were used: the colon, kidney, and skin (US Biomax, MD). One ROI was chosen for each sample. The three tissue samples were scanned and measured by the four imaging devices to generate 12 images.

4 | RESULTS

Samples of spectral transmittance of the tissue slides are shown in Figure 4. Figure 4A shows the spectra of five locations of the colon tissue sample. Spectrum #1 is a void area whose spectral transmittance curve is a flat line at nearly 100%. Spectra #2 and #3 are pink shades resulted from the eosin stain. Their spectral transmittance curves are similar but different in magnitude. Although these shades are frequently described as pink in H&E stain images, according to their spectra, which seem to comprise red and blue components, light purple might be a better descriptor than pink. Spectra #4 and #5 are darker purple shades resulted from the hematoxylin stain. Figure 4B shows the spectra of four locations of the kidney tissue sample. Spectra #1 (white), #2 and #3 (pink), and #4 (purple) exhibit similar curves to those in Figure 4A. The kidney tissue sample has more void area than the colon. Figure 4C shows the spectra of four locations of the skin tissue sample. Spectra #1 (white), #2 (pink), and #4 (purple) resemble those in (a) and (b), except for spectrum #3 (brown), which is part of a hair. Nearly a quarter of the skin tissue sample is void.

The 12 images generated by the four imaging devices from the three tissue samples are shown in Figure 5. The

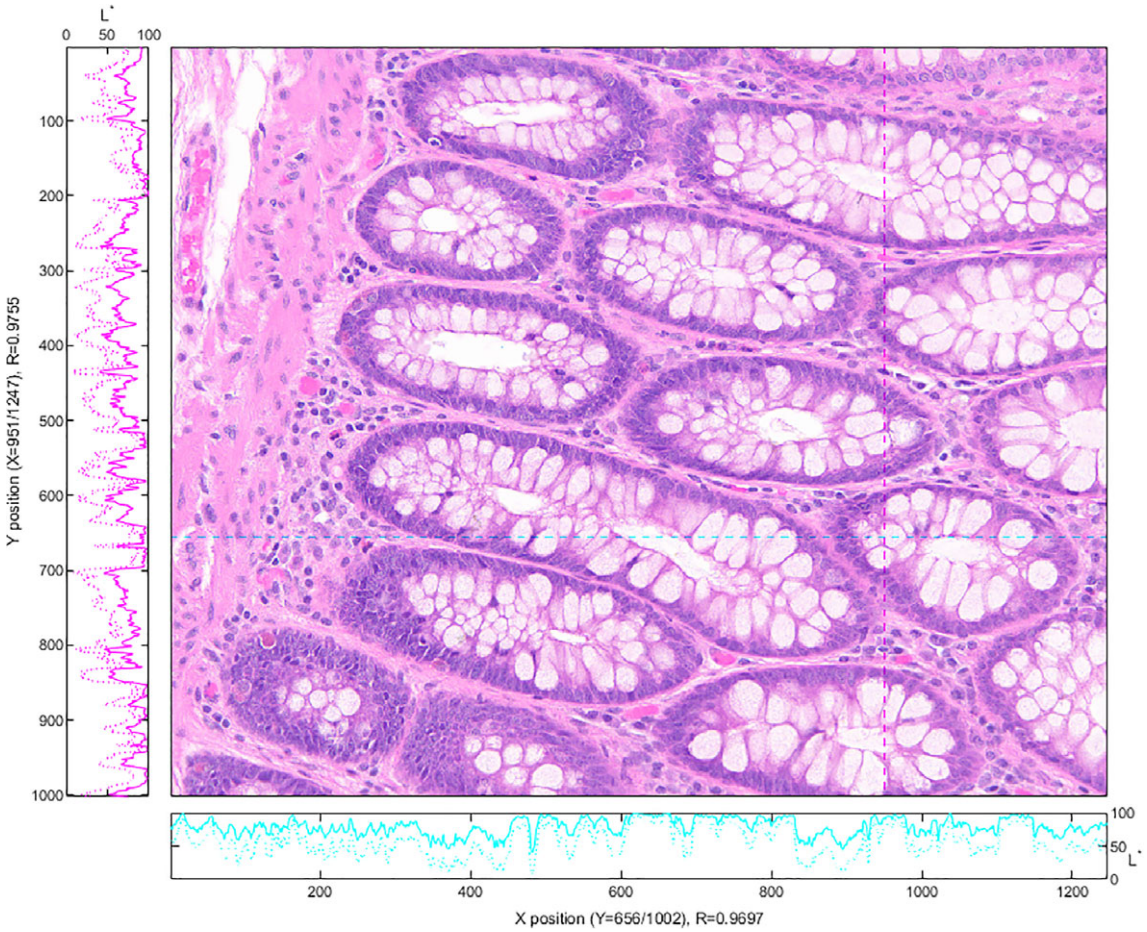

FIGURE 3 Correlation analysis of the registered images between the truth and the modern device. The correlation was calculated based on CIE L^*, while the reconstructed truth image (sRGB) is shown to visualize the image content. The cursor was used to choose vertical (pink curves, $x = 951$) and horizontal (cyan curves, $y = 656$) cross sections. The CIE L^* values of the cross sections are shown beside the x- and y-axes. The solid curves represent the truth and the dotted curves of the modern device, which generated lower L^* values overall. The correlation coefficients are 0.9755 and 0.9697 for vertical and horizontal cross sections, respectively

colon, kidney, and skin tissue images are on individual rows from top to bottom. The images generated by the modern and legacy WSI scanners are in the second and third columns, respectively. The images in the first column, called the truth images, are reconstructed from the spectral data measured by the multispectral imaging system. The last column contains monochrome images that were derived from the truth images.

Despite the potential color errors occurring in the reproduction media, it is identifiable that the legacy WSI scanner (third column) produced color shades that give a hint of overstaining and deviate from both the truth and the modern WSI scanner. In contrast, the monochrome images (fourth column) bear no chromaticity and look less informational than the rest of the images.

The quantitative color differences of the two commercial WSI scanners were calculated as shown in Figure 6, where the CIEDE2000 of each pixel is represented in height and pseudocolor to replace the original pixel. Although these plots represent the color differences, the terrains in all 6 three-dimensional plots resemble the original structure. In all plots, the void areas on the glass slides have lower color differences, so both plots of the skin tissue sample (third row) have a large portion of blue data points near zero. The terrains in the three legacy WSI images (second column) have higher elevations. The outlines or boundaries of different tissues seem to generate greater color differences. These plots offer a tool for pinpointing the color differences of a certain tissue, cell, or pixel of interest.

Figure 7 shows the histograms of the color differences in the nine images from the modern, legacy, and monochrome WSI scanners. For all three tissue samples, the maximum color differences of the legacy WSI scanner (dashed curves) were greater than 40 ΔE_{00}, while those of the modern WSI scanner (solid curves) were less than 30 ΔE_{00}.

For the colon and kidney tissues (left and center plots), the ΔE_{00} peaks were in the ranges of tens and twenties for the legacy and modern WSI scanners, respectively.

For the skin tissue (right plot), the void areas on the glass slide caused color differences related to white balance, which created very narrow and tall spikes at about 5 ΔE_{00}. Neglecting these spikes, the peaks were 8 ΔE_{00} and 27 ΔE_{00} for the modern and legacy WSI scanners, respectively. The minimum color differences cut off at about 4 ΔE_{00}.

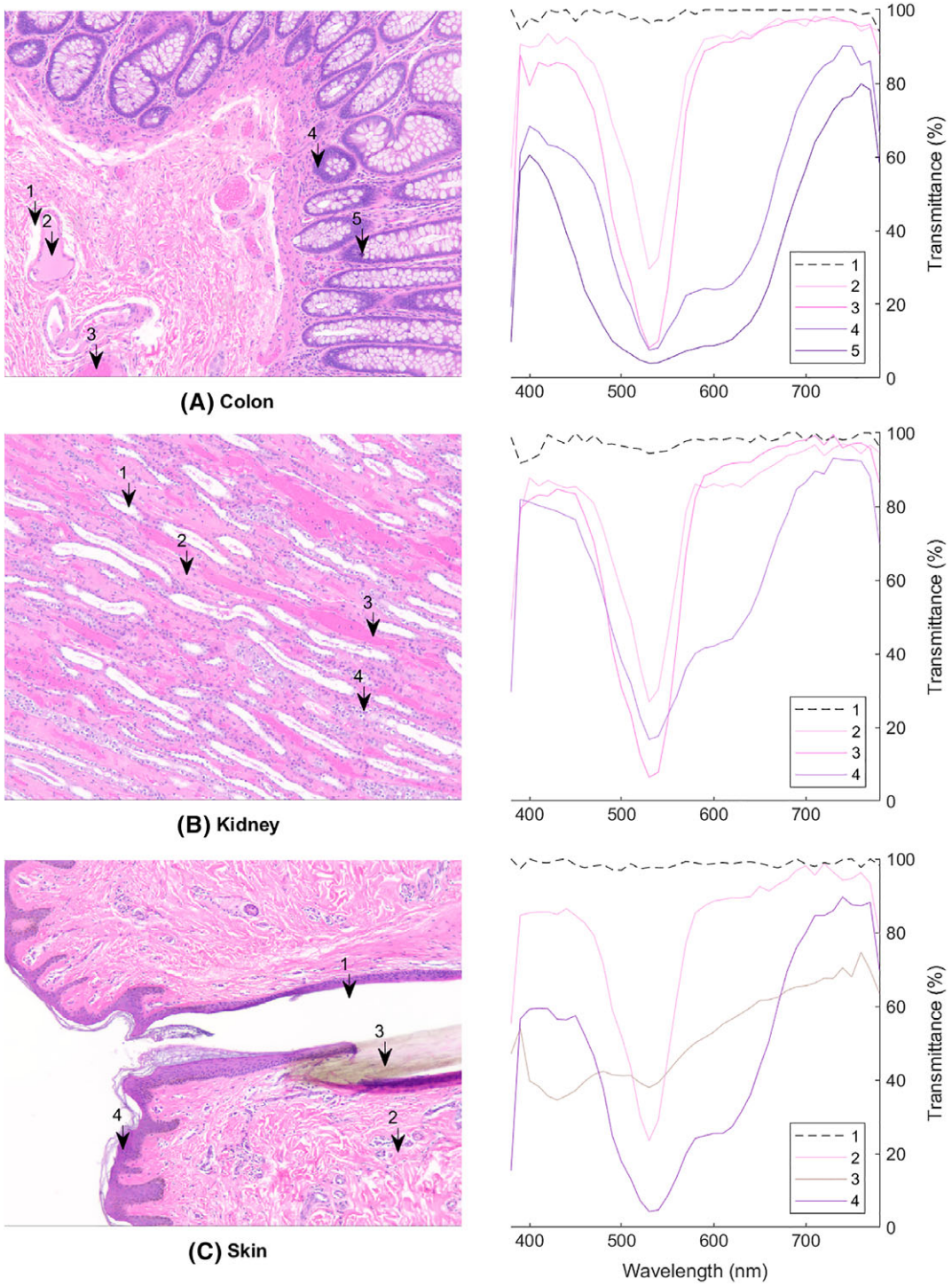

FIGURE 4 Spectral transmittance samples (right column) of the colon (top row), kidney (center row), and skin tissue samples (left column). Each curve represents the spectrum of a single pixel in the image

Figure 7 also shows the data of the monochrome images (dotted curves) for comparison. The data formed tub-shaped curves with two peaks for all three tissues. The right-side peaks were located about the same as or slightly to the right of those of the legacy WSI scanner (dashed curves), indicating that a major portion of the monochrome pixels carried equal or slightly greater than most pixels from the legacy WSI scanner. The left-side peaks were very tall and narrow, located at exactly zero ΔE_{00}. These pixels were in the void areas on the glass slide and immune from the white balance issue because they were converted from the truth image directly. Also note the dense population distributed between 0 and 7 ΔE_{00}, where the modern and legacy WSI scanners had very few.

Table 1 lists the mean and SD of the color differences with respect to the truth images. For all three WSI scanners, the colon slide resulted in the greatest color differences followed by kidney and then skin. Overall, the legacy WSI

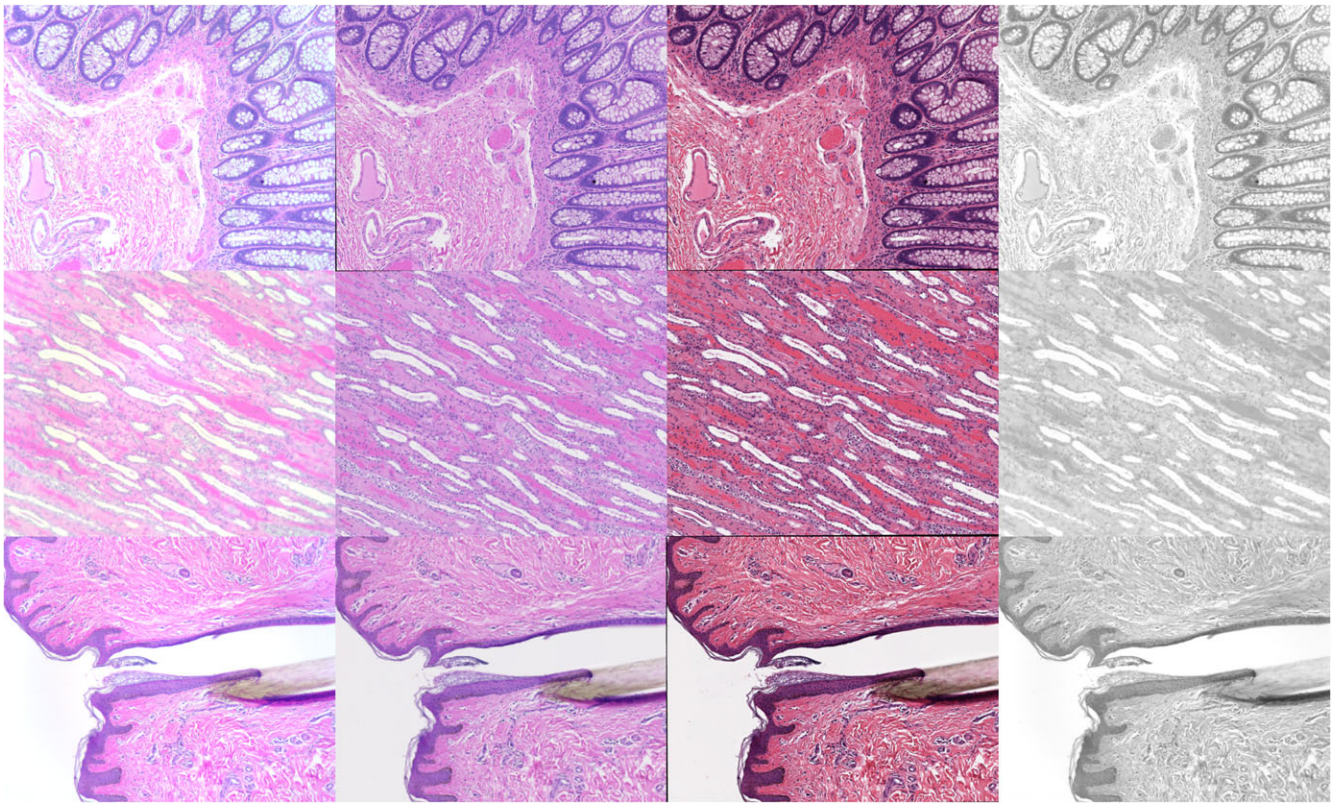

FIGURE 5 Images obtained in this study. From left to right, the columns contain images from the multispectral imaging system, modern, legacy, and monochrome whole-slide imaging (WSI) scanners. From top to bottom, the rows are for the colon, kidney, and skin tissue samples

scanner generated greater color differences in a wider range compared with the modern one. For all three tissue types (colon/kidney/skin), the mean color difference of the legacy WSI device is about twice that of the modern WSI device (13.08/11.33/8.40 vs 26.33/22.69/18.02). The standard deviations are also doubled (3.94/3.91/4.73 vs 7.88/8.70/10.41).

The mean color errors in ΔE_{00} are in the range of [8.40,13.08], [18.02,26.33], and [17.42,17.62] for the modern, legacy, and monochrome WSI scanners, respectively. Compared with the monochrome images, the legacy WSI scanner actually produced greater mean color differences for all three tissue samples (26.33/22.69/18.02 vs 17.62/17.54/17.42). In other words, in terms of minimizing the color differences from the truth, the hypothetical monochrome device outperformed the legacy WSI scanner. The finding of a monochrome device outperforming a color device is somehow counterintuitive and needs to be interpreted thoughtfully. The only color difference between the truth and the monochrome image is the chromaticity. The legacy device generated a color difference greater than the chromaticity of the truth because both the hue and chroma were exaggerated as observed in Figure 5. The legacy device was outperformed not because its images carry less color information than the monochrome images, but because its brightness/color contrast was overenhanced. As a result, the monochrome images may look more faithful in terms of color reproducibility than the color images. Although the monochrome device outperformed the legacy one, it was outperformed by the modern one for all three tissue samples (13.08/11.33/8.40 vs 17.62/17.54/17.42).

To compare relative performance of the commercial WSI scanners with respect to the hypothetical monochrome device, Table 2 lists the normalized mean color differences. Compared with the monochrome device, the modern WSI scanner reduced the color differences to 74%, 65%, and 48% for the colon, kidney, and skin tissue samples, respectively. In other words, in the best-case scenario, the skin tissue sample, the modern WSI scanner cut only 52% color differences compared with a monochrome device. Considering either the absolute ΔE_{00} (8.4-13.08) or the normalized percentage (48%-74%), there seems to be room for the modern WSI scanner to improve.

To investigate the scenarios in which the monochrome device outperformed the commercial WSI scanners, the color differences were visualized as shown in Figure 8. The images generated by the commercial WSI scanners were augmented by replacing every pixel that has a less color difference than the monochrome device with a black pixel. Thus, the remaining nonblack pixels represent the pixels that were outperformed. The area (or percent) of outperformed pixels is annotated in the image label. For the colon tissue sample, 85.31% of the legacy images were outperformed by the monochrome device except for some pink shades, while the modern WSI scanner had 28.19%, which consists of mostly bright purple shades. For the kidney tissue sample,

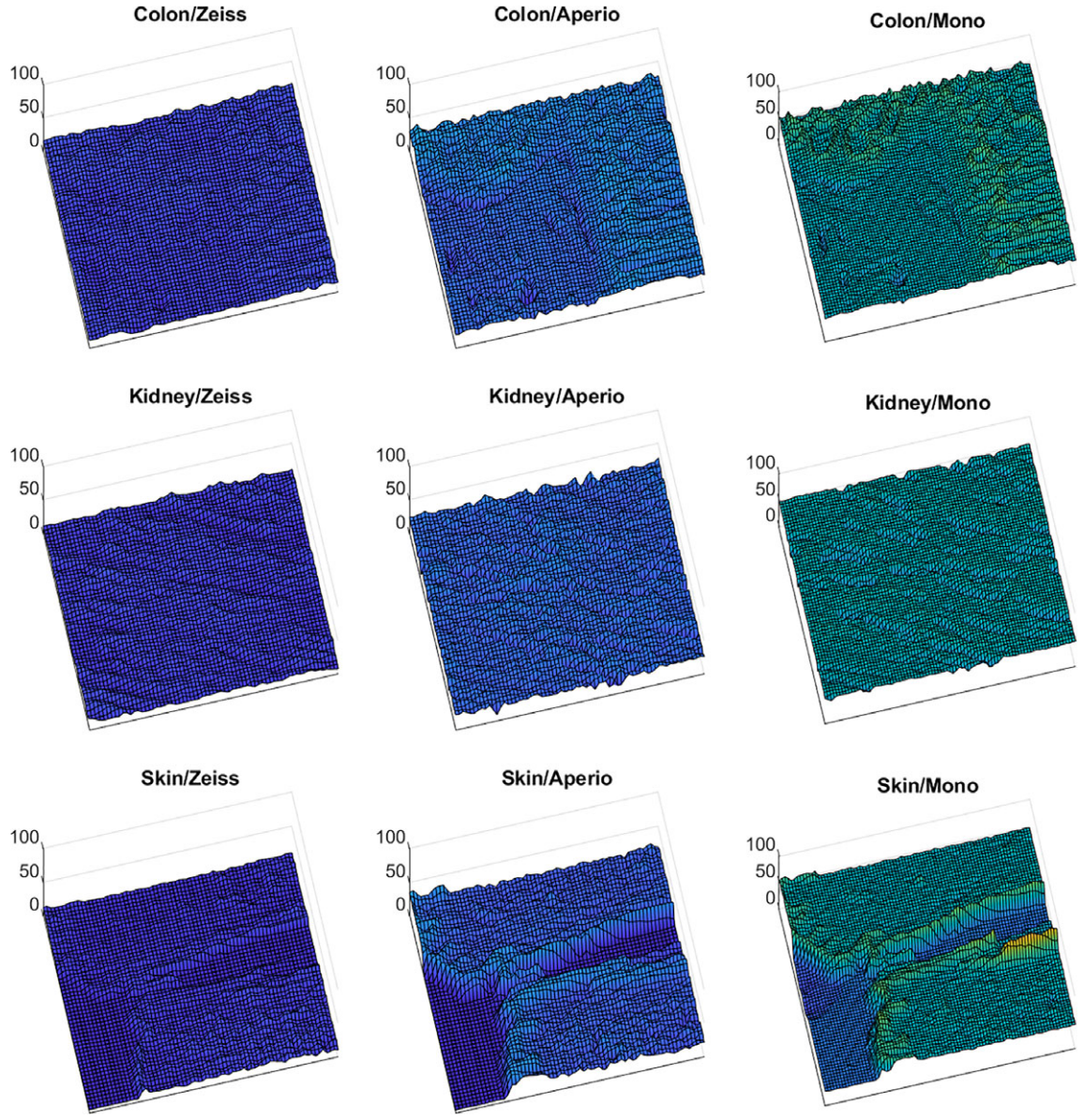

FIGURE 6 ΔE_{00} for each pixel shown in three-dimensional heat maps

the legacy WSI scanner generated 63.98% outperformed pixels, which consist of both pink and purple shades and void areas. The modern WSI scanner had 25%, which are mostly void areas and bright purple shades. For the skin tissue sample, the void areas constitute the major portion of the outperformed pixels for both WSI scanners. The purple

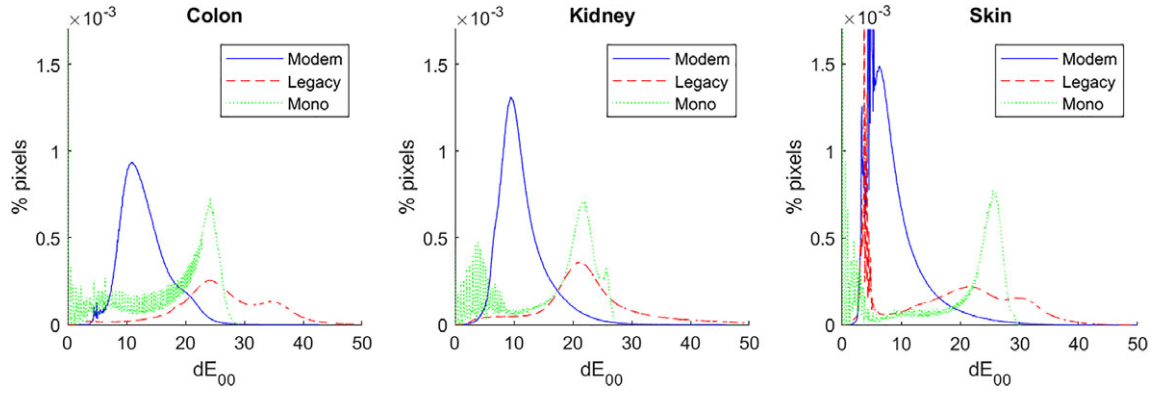

FIGURE 7 Histograms of ΔE_{00} for three tissue samples

TABLE 1 The mean and SD of color differences ΔE_{00}

	Modern		Legacy		Monochrome	
Tissue	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Colon	13.08	3.94	26.33	7.88	17.62	8.06
Kidney	11.33	3.91	22.69	8.70	17.54	7.12
Skin	8.40	4.73	18.02	10.41	17.42	10.20

TABLE 2 Normalized mean ΔE_{00} with respect to the monochrome images

Tissue	Modern (%)	Legacy (%)	Monochrome (%)
Colon	74	149	100
Kidney	65	129	100
Skin	48	103	100

shades contributed the second largest portion for the legacy WSI scanner.

5 | DISCUSSION

It is well documented that the eosin dye in the H&E stain is autofluorescent, indicating that it absorbs light at a certain wavelength and emits light at another wavelength. The difference between the exciting wavelength and the emitted wavelength might introduce errors when estimating the color truth with a multispectral imaging method depending on the measurement geometry. It is noticed that, in this study, the measurement setup was transmissive rather than reflective.

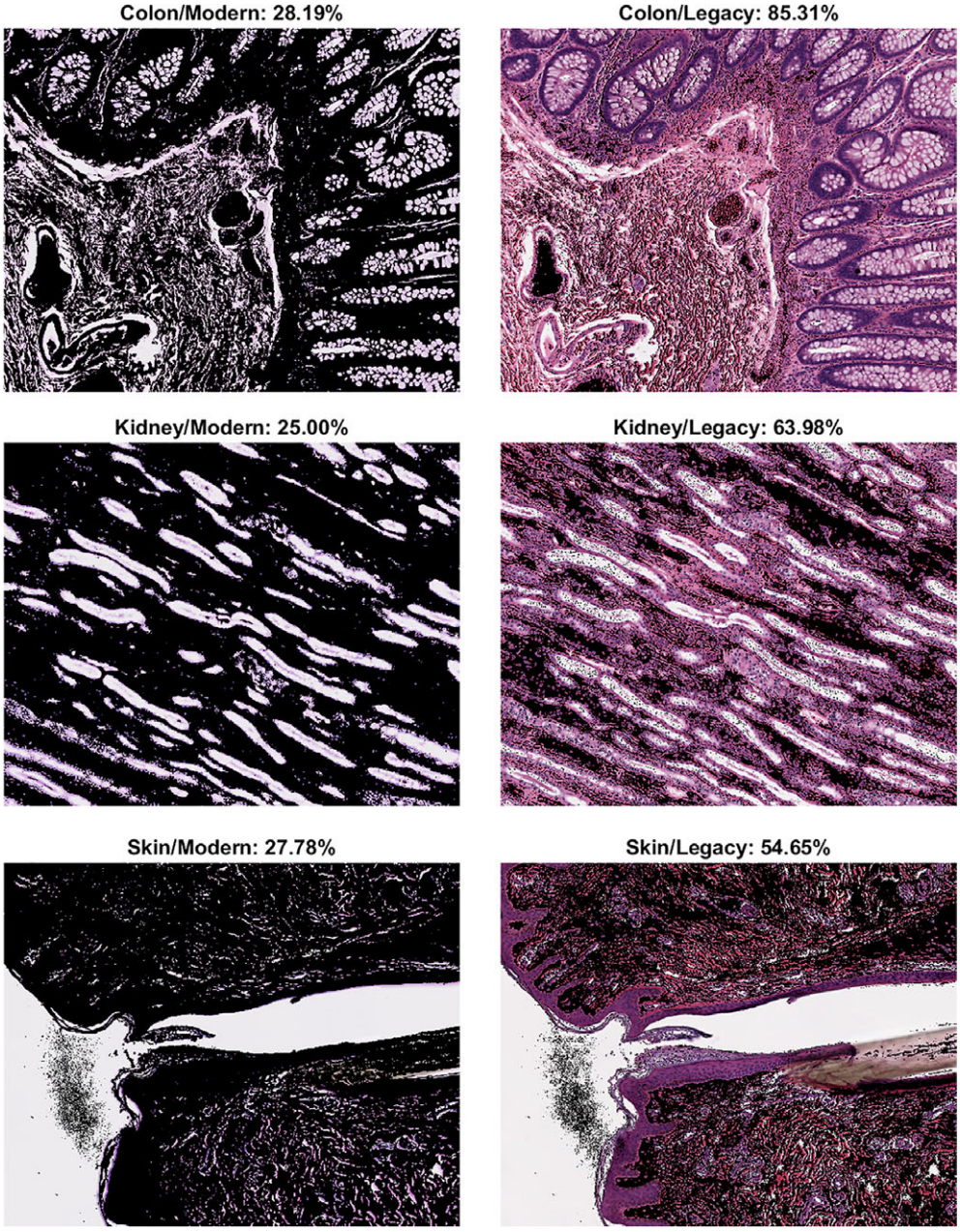

FIGURE 8 The pixels that have higher ΔE_{00} than the monochrome device. Augmented images showing colored pixels that have higher ΔE_{00} than their counterparts from the monochrome device of the same tissue type. The remaining pixels, which have lower ΔE_{00}, were filled with black

In our experiments, the transmitted light completely dominated the reading so the autofluorescence of eosin was unmeasurable, which was confirmed by another verification test using a spectroradiometer to measure the response from 380 to 780 nm at the interval of 10 nm. Therefore, the eosin fluorescence can be ignored without affecting the measurement accuracy.

The eosin is also known for fading when exposed to light. In the study, the H&E glass slide was scanned three times (by the multispectral imaging system, the legacy WSI device, and the modern WSI device) only within several hours. Therefore, it was very unlikely to observe any measurable difference between scans caused by eosin fading. Like any color targets or stained histological samples, the H&E glass slides should be kept in a cool and dark place between experiments. Target calibration, that is, obtaining the color truth with the multispectral imaging system, may be required when the H&E glass slide is reused after an extended period of time or extensive use under intense illumination.

In this study, a tunable light source was used to separate the light into different bands. Compared with conventional multispectral imaging methods that added narrowband filters onto the light path of an existing microscope system and greatly reduced the illumination, the tunable light source used in this study was powerful enough so that the camera sensor noise was not noticeable. For instance, at 550 nm, the measured luminance of the narrowband light was greater than 90 cd/m^2.

The essential method of this study, multispectral imaging, is a well-developed imaging technique. Multispectral imaging can be implemented in various ways depending on the required accuracy and available resources. All the hardware components used in this study are commercial off-the-shelf parts, so the whole system can be duplicated. Alternatively, commercial instruments or measurement services are available for users who require the spectral transmittance of tissue slides.

Multispectral imaging enabled the method of using biological tissue slides to test the WSI scanners, so that the assessment results, compared with man-made color targets, were more relevant to real-world individual cases. Nevertheless, the biological tissue slides cannot be duplicated, so comparing two WSI scanners that are remotely located with the same tissue slides becomes inconvenient due to the transportation of either the slide or the scanners.

In this study, only H&E stained glass slides were used. Although H&E is the most commonly used stain, its color gamut is mostly limited to shades of light purple and dark purple. Other stains resulting in different color gamut should be included in the future work.

The CIEDE2000 formula was used to calculate the color difference. Although it is a well-accepted measure in the color research community for predicting perceptual color differences, the interpretation of the numerical ΔE_{00} may not seem to be intuitive for many WSI scanner users. In this study, the monochrome device was included as a reference to assist the interpretation of the ΔE_{00} results. By referencing to the truth image and the monochrome image, namely the color truth and the color falsity, the experimental data show that the legacy WSI scanner was outperformed by the monochrome device, while the modern WSI scanner performed halfway between the truth and the falsity.

Although the CIEDE2000 formula was used in this study, the other color difference formulas should also be considered based on the purposes and the scanner performance. CIE TC 1.57 recommends using CIEDE2000 for errors up to five CIELAB units and CIE76 for errors greater than five.[14] In this study, CIEDE2000 was chosen in the protocol before the experiments were conducted. In the future work, various color difference formulas, including CIE76, CIE94, and CIEDE2000, should all be considered in the evaluation.

This study did not assess the counterpart of the WSI scanner, the display subsystem, which is equally important. A study combining both subsystems should be included in future work.

This study was designed based on the assumption that the rendering intent of the WSI scanners is to reproduce the original color of the scene. In other words, color reproducibility is the objective of the assessment. However, the rendering intent of a color imaging device does not need to be color reproducibility. For example, the rendering intent of mobile phones may not be faithfulness but esthetics. To the best of our knowledge, none of the WSI systems marketed in the United States, including the two commercial systems tested in this study, explicitly states the color rendering intent in the documentation. Without defining the color rendering intent, a color imaging device cannot be assessed colorimetrically and compared with another device. For example, some physicians may prefer the pronounced brightness/color contrast of the legacy WSI scanner to the truth. When it comes to evaluating two WSI scanners that both produce pronounced color contrast but render color very differently, while neither of their rendering intents was defined, an objective conclusion will be very challenging to reach until a costly reader study is commissioned. The ultimate goal of this study is to advocate quantitative assessment methods for assessing the color performance of medical imaging devices. Quantitative assessment methods will assure the safety and effectiveness of medical imaging devices and provide science-based information for patients, caregivers, and providers.

6 | CONCLUSION

The color truth of the glass slide is required for assessing the color performance of WSI scanners and can be a challenging task without proper instrumentation. A multispectral imaging system was developed in this study to measure the color

truth of biological tissue samples and eliminated the need of using man-made color targets, which do not resemble the color gamut and the microscopic structure of real tissues. The multispectral imaging system measured the spectral transmittance of the tissue samples at the pixel level, so the color error of each pixel in the whole-slide images could be calculated and analyzed. Three biological tissue samples, including the human colon, kidney, and skin, exhibiting variation in color gamut and microscopic structures were used to test two commercial WSI scanners that were manufactured 6 years apart. Besides the color truth, the monochrome image was also derived from the color truth to serve as a reference representing a standardized poor performer. The experimental results showed that the legacy WSI scanner was outperformed by the monochrome device, while the modern WSI scanner cut the color differences of the monochrome device by 52% at most. The findings suggested that there is still room for the WSI scanners to improve.

ACKNOWLEDGMENTS

The authors thank the reviewers for their valuable comments. This study was supported in part by a grant from the Critical Path Initiative at the Food and Drug Administration and by an appointment to the Research Participation Program at CDRH administered by ORISE through an interagency agreement between DOE and FDA. The mention of commercial products herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

ORCID

Wei-Chung Cheng https://orcid.org/0000-0001-5259-2990

REFERENCES

[1] Badano A, Revie C, Casertano A, et al. Consistency and standardization of color in medical imaging: a consensus report. *J Digit Imaging*. 2015;28(1):41-52.

[2] Abels E, Pantanowitz L. Current state of the regulatory trajectory for whole slide imaging devices in the USA. *J Pathol Inform*. 2017;8:23.

[3] Parwani AV, Hassell L, Glassy E, Pantanowitz L. Regulatory barriers surrounding the use of whole slide imaging in The United States of America. *J Pathol Inform*. 2014;5:38.

[4] FDA. *Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices—Guidance for Industry and Food and Drug Administration Staff*. Silver Spring: Hampshire Ave; 2016.

[5] Cheng W-C, Keay T, O'Flaherty N, et al. Assessing color reproducibility of whole-slide imaging scanners. *Medical Imaging 2013: Digital Pathology*, Vol. 8676. International Society for Optics and Photonics; 2013.

[6] Shrestha P, Hulsken B. Color accuracy and reproducibility in whole slide imaging scanners. *J Med Imaging*. 2014;1(2):027501-027501.

[7] Wei H, Brill MH, Park T. Evaluation of targets for color calibrating digital images from an optical bright-field transmission microscope. *Color Res Appl*. 2015;40:577-584.

[8] Yagi, Y. Color standardization and optimization in whole slide imaging. *Diagn Pathol*. 2011;6(1 Suppl):S15.

[9] Clarke EL, Revie C, Brettle D, et al. Development of a novel tissue-mimicking color calibration slide for digital microscopy. *Color Res Appl*. 2017;43:184-197.

[10] Saleheen F, Badano A, Cheng W-C. Evaluating color performance of whole-slide imaging devices by multispectral-imaging of biological tissues. *Medical Imaging 2017: Digital Pathology*, Vol. 10140. International Society for Optics and Photonics; 2017.

[11] Fischer C, Kakoulli I. Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Stud Conserv*. 2006;51(suppl 1):3-16.

[12] IEC. *IEC 61966-2-1: Multimedia Systems and Equipment-Colour Measurement and Management-Part 2-1: Colour Management-Default RGB Colour Space-sRGB*. Geneva, Switzerland: International Electrotechnical Commission; 1999.

[13] Sharma G, Wu WC, Dalal EN. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl*. 2005;30(1):21-30.

[14] Green P, MacDonald L, eds. *Colour Engineering: Achieving Device Independent Colour*. Chichester, England: John Wiley & Sons; 2011.

AUTHOR BIOGRAPHIES

W. CHENG is a research scientist at Food and Drug Administration (FDA). He received a PhD in electrical engineering from University of Southern California. His research interests include color imaging systems, color vision, and display systems.

F. SALEHEEN is currently a research scientist at EDDA Technology, Princeton, NJ. He received PhD and MS degrees in electrical engineering in 2017 and 2014 from Temple University, and obtained a BS in electrical and electronic engineering in 2008 from the Bangladesh University of Engineering and Technology. In 2016, he participated in FDA's research project on color performance assessment of whole-slide imaging device for digital pathology as an ORISE research fellow. His research interests include medical robotics, statistical and game-theoretic control, multispectral imaging, and tactile sensing.

A. BADANO is a senior biomedical researcher at FDA. He currently serves as Deputy Director of the Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health. He received an MEng degree in radiological health engineering and a PhD degree in nuclear engineering from the University of Michigan in 1999 and 1995 after obtaining a Chem Eng degree from the Universidad de la República, Montevideo, Uruguay, in 1992. His primary interests are in the characterization and modeling of medical imaging acquisition and visualization systems.

How to cite this article: Cheng W-C, Saleheen F, Badano A. Assessing color performance of whole-slide imaging scanners for digital pathology. *Color Res Appl*. 2019;44:322–334. https://doi.org/10.1002/col.22365