A Novel, Quantitative, Automated Technique for Identifying and Scoring Stained Regions in Immunohistochemical Specimens with Dual Stains

by: R. Wade Schuette
University of Michigan
Ann Arbor, Michigan
schuette@umich.edu

V1.1, Posted December 24, 1998 to web site, and publicized on NIH-Image website

Dedicated to the memory of Bob Stafford and all the members of the Breast Cancer List whose courage inspires renewed energy in the search for a cure.

Abstract

Reliable, reproducible quantitative scoring of immunohistochemically stained images has been difficult, due in part to the fact that the stain and counterstains in common use have overlapping colors. This paper presents the mathematical basis for completely and reproducibly identifying the amount of each stain at each pixel of the image, regardless of image depth and overlapping stains. The technique allows closed-form reconstruction of an image of each stain alone, which can be quantified, or ratio of the two stains, which produces a substantially higher signal-to-noise ratio than other techniques now in use. The technique is only valid for images with two different stains. Preliminary empirical investigation is encouraging that this technique can be used in practice.

1. Introduction -- the problem

The increase in use of image processing techniques and image-based measures, such as the Ki-67 proliferation index, has been plagued with a wide variability between studies as well as inter- and intra-reader variability. As the use of color images and specimens stained with multiple color stains increases, some of the sources of variability need to be addressed.

Some variations, such as changes in brightness across the image, can be corrected by proper calibration of the microscope, although in practice this step is too often neglected. A simple tour of the web shows many images with obvious darkened corners and even color gradients across the background, or obvious artifacts due to excessive compression.

Other variations, such as lamp color-temperature, are more insidious. Even careful operators who set the lamp exactly on, say, “9” for each image across a long study, forget that the line-voltage may easily vary by 10% or more in most laboratory buildings. Many operators simply turn up the lamp voltage until the image is just short of saturation -- a brightness that may vary with specimen thickness or convenience.

The problem is that tungsten filaments obey a law such that the proportion of colors in the emitted light depends dramatically on the voltage / brightness. A 10 percent variation in voltage will result in a 20% variation in power and filament temperature.

It's obvious on inspection that at low voltages the light appears somewhat more yellow than at high voltages, but it's not so obvious that the ratio of red to blue varies by a factor of 3 between low lamp voltages (2000 degrees K) and high lamp voltages (3200 degrees K).

The implication of that subtle factor is that seemingly robust measures, such as using the ratio of red to blue as a "cutoff" for counting a feature, can vary also by a factor of 3 due to lamp voltage.

In some labs, expensive equipment is used to regulate the lamp voltage or brightness exactly, but it turns out that even this is subject to problems due to variability in specimen thickness over the lifetime of a study. Most operators are unaware that a setting of "4" on a common microtome results in a slice thickness of 8 microns, or twice the setting. Specimens as thick as 12 microns have been measured by the author with a confocal microscope on slices nominally at "4" micron thickness. Across a multi-year study with different persons preparing slides, some variability is unavoidable, but measuring every slide is prohibitively expensive.

The problem is that the darkness of a specimen varies non-linearly, and differently for blue and red. Suppose for example that brightness is given by

Red Brightness = 8 * exp (- c1 * T)
Blue Brightness = 6 * exp (- c2 * T)

such that at a given thickness T (say 4 microns), these are equal. Suppose exp(-c1*T) = 0.6 of full brightness, and exp (-c2 * T) equals 0.8 of full brightness. These two would be considered "equal" since 8*0.6 = 6 * 0.8 = 4.8.

Suppose the thickness of the next slide, of identical tissue, is twice as thick. The red brightness will be 8*0.6*0.6 = 2.88. The blue brightness will be 6 * 0.8 * 0.8 or 3.84, no longer equal. The little appreciated fact is that the ratio of red to blue will vary with slide thickness, and dramatically.

With a possible net factor of 3 due to color temperature, and another factor of 3 due to slice thickness, we already have a possible factor of 9 times variation in identical specimens. Little wonder that studies tend to be non-reproducible and not comparable when color imaging is used, even carefully. This increased noise means that much larger study sizes are required to accomplish the same statistical power.

The author's experience with such problems led to the search for a measurement algorithm that was independent of color-temperature and slice thickness entirely, so that the same slide, measured at different sites on different days by different operators, would yield the same score for fraction stained, percent more red than blue, etc.

The "V-algorithm", below, provides such an algorithm. This paper explains the theory, and accompanying materials provide examples. The hope is that a much stabler measure will not only save time and money by reducing necessary sample sizes, but will also allow meaningful comparisons across sites and studies, making the entire study more valuable as well.

A surprising side benefit is that the actual amount of each stain can be quantitatively captured even for regions where the stains overlap substantially. allowing a greater proportion of each slide to be used and reducing selection bias.

2. The "V" Algorithm

Stains affect the light passing through them by absorbing some fraction of the incoming light, and transmitting the rest. The effect is complementary -- if the stain absorbs more green light than red light, then the transmitted light will appear somewhat red in hue, and will appear somewhat darker than the original light source.

If a second slide of the same thickness and color is added to the light-path, the impact will be to produce an even darker light, that is even redder than one slide alone would produce.

Mathematically, the situation can be described as an intensity which is reduced exponentially as the amount of stain in the light path increases. If the amount of stain is "x", and the source light intensity is captured as a brightness value of S the brightness of that pixel would be given by

(1) $$\text{Brightness} = S * \exp(-x/C)$$

where C is a constant that can be fitted empirically to a given stain.
Equivalently, taking the natural log (ln) of each side of that equation, we have

(2) $$\ln(\text{brightness}) = \ln(S) - (x/C)$$

which is the equation of a straight line with intercept ln(s) and slope -1/C.

The constant "C" is different for red light than for blue light, and each stain has a different, but fixed constant. (An assumption that needs to be empirically confirmed and documented.) If red and green channels are not calibrated to equal maximum brightness, we have a maximum brightness for red "Sr" and for green "Sg".

If we define Car as the constant for stain "a" absorbing red light "r", and similarly define Cbr as the constant for stain "b" absorbing red light, Cag the constant for stain a, green light, and Cbg the constant for stain b, green light, then the observed brightness of a given pixel which has a light beam going through amount Xa of stain A, and amount Xb of stain B, would be given as follows:

(3.a) Red channel: $$\text{brightness}(Xa,Xb) = Sr * \exp(-Xa/Car) * \exp(-Xb/Cbr)$$
(3.b) Green channel: $$\text{brightness}(Xa,Xb) = Sg * \exp(-Xa/Cag) * \exp(-Xb/Cbg)$$

or, equivalently, taking the natural log of each equation,

(4.a) Red channel: $$\ln(B(Xa,Xb)) = \ln(Sr) - Xa/Car - Xb/Cbr$$
(4.b) Green channel: $$\ln(B(Xa, Xb)) = \ln(Sg) - Xa/Cag - Xb/Cbg$$

We can see that, if the constants C were known, the Source brightness S could be empirically measured (a blank spot on the slide), the left side brightnesses for a given pixel are measured by the camera/image, and we're left with two equations in two unknowns (the amount Xa of stain A and Xb of stain B in the light path) which can be solved for in closed form.

In other words, if we knew the constants C, we would only need to measure the brightness of an empty slide to get S, and then from observed values of the Red and Green channels for a given pixel, we could unambiguously extract

the amount of stain A and stain B in the light path.

Or, if we knew the constants to within a constant scale factor for all of them, we could still unambiguously extract the ratio of stain A to stain B for any pixel.

Fortunately, the constants can be measured post-hoc for dual-stained images, as explained next.

If we plot on a graph the log of the Green value of each pixel vs the log of the Red value, we get a graph that looks like **figure 1** below.

The "white point" corresponds to zero stain in the light path for either of the two channels being plotted. That is on the upper right end of the inverted "V".

The top edge of the "V", shown in green below, corresponds to colors possible for various amounts of only stain A. No stain A corresponds to the white point, and as the amount of stain A increases, the resultant color moves down and to the left along that straight line.

Similarly, the bottom edge of the "V", shown in red below, corresponds to color pairs possible for various amounts of stain B, increasing down and to the left.

Any possible pair of Red and Green values, such as point "P" in the figure, corresponds to a unique amount of stain A (Xa in the figure) and a unique amount of stain B (Xb in the figure.)

Note that, in this ideal model, the only possible pixel values occur in the green shaded area inside the "V".

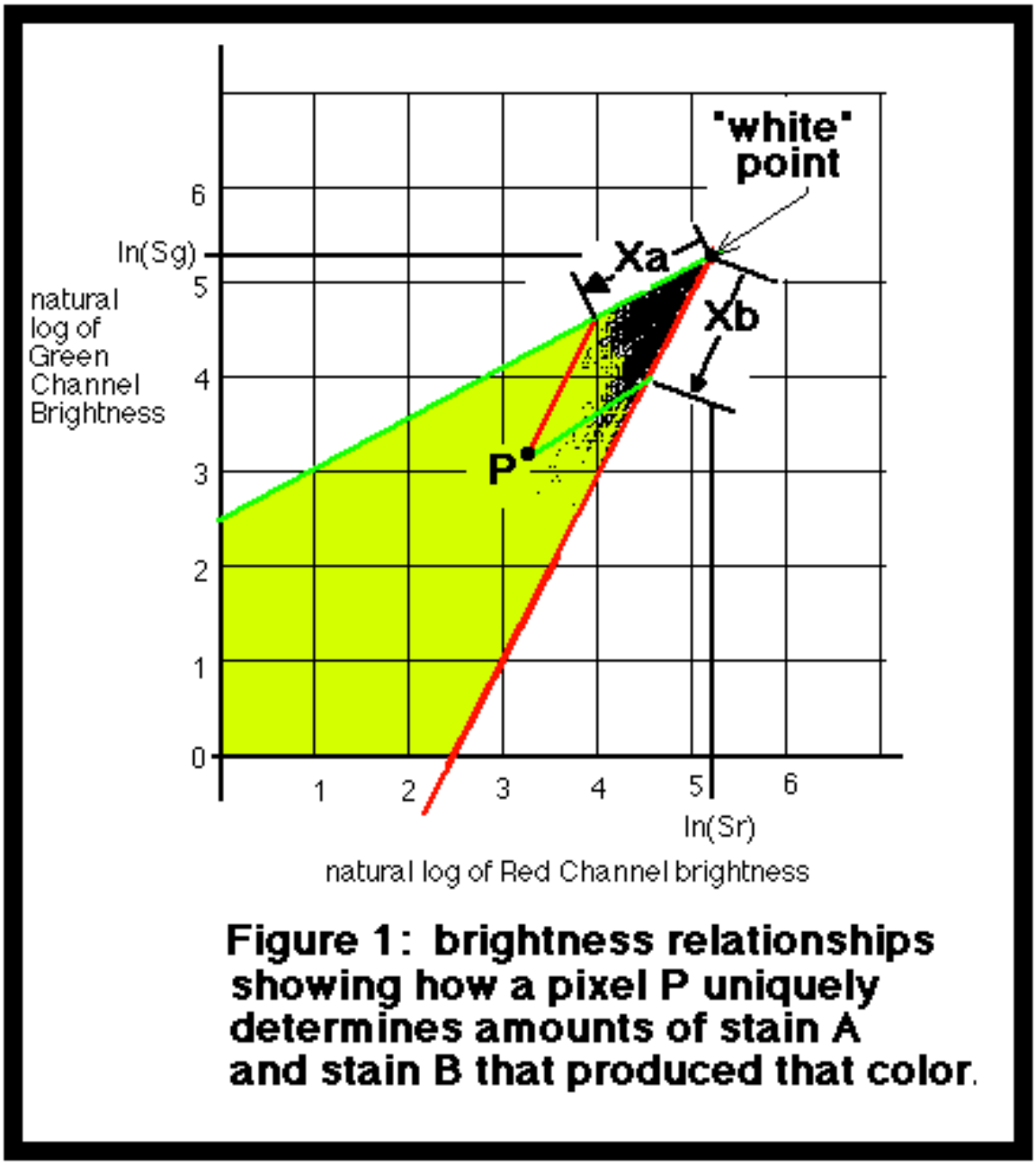

Figure 1: brightness relationships showing how a pixel P uniquely determines amounts of stain A and stain B that produced that color.

Furthermore, for most images, there will be sufficiently dense concentration of points near the top right edge that the two edges of the "V" can be determined quite accurately by inspection. For automated techniques, this fitting process would, of course, be automated to maximize the statistical value of each of the pixels available.

3) Summary of the algorithm

In summary then, the steps to separating the stains are as follows:

1) any two color channels are chosen, say green and red.

2) Every pixel pair in the original image with red, green, blue components (R,G,B)

is plotted on one scatterplot as in figure , plotting ln(G) vs ln(R).

3) The scatterplot should form a "V" shape, with vertex at the "white point", which will be at red value ln(Sr) and green value ln(Sg). In practice, the absolute brightest points are often absent, but the sides of the "V" can be determined and extended to locate the white point.

4) From the white point and the slope, or equivalently from any two points on one edge of the "V", a pair of unknown coefficients C can be determined. From both edges, all four of the coefficients C in equations 3 and 4 can be empirically determined for that image and camera and capture settings.

5) Equations 4 can then be solved, for each (R,G,B) pixel observed, to give the relative amounts of stain A (Xa) and stain B (Xb) that are the only solution to equation 4 giving that RGB triplet.

6) The result is an image map of each stain , that can the be quantified directly, masked, ratioed, or whatever.

4) Discussion and benefits of this algorithm

All the above assumes that the image has been "flat-fielded", i.e., that the intensity of light is equal at all points on the image. If the image has varying brightness behind it, the "V" shape will be diffused.

Note that it is possible to determine the white point for a sub-image even if there is no "clear" spot, by extrapolating the edges of the "V" back to the vertex. This works regardless of the color temperature of the source, and does not require Sr = Sb = Sg. So long as the camera has a linear mapping of brightness to pixel value, two images taken under different brightness and source color temperatures should still yield exactly the same stain images for Xa and Xb.

The constants for given stains may be constant for a given type of image, and, in fact, independent of camera gain. A higher gain would expand the pixel array down and to the left, but the edges would still describe the exact same lines. This means that the ratio of stains can be measured independently of the camera gain setting, valid even if two slides have different thickness specimens on them.

Conversely, if a set of images with known absolute concentrations of stain (by some other means) is used for calibration, an exact quantitative relationship can be established between pixel positions and, say, micrograms of stain in the light path, for a given camera gain setting.

A side benefit of this technique is that it unambiguously identifies the "white point", which could be reset to full brightness (say 255) by multiplying pixel values by some constant, allowing a sort of standardization of images (say H&E's) with a standard defensible color-balance, that would overcome the wide range of color-balances observed in published images today.

5) Future Work

The next step in investigation of this technique is to do an empirical study of how well it works in practice.

The author has written code that implements this algorithm in NIH Image and investigated several H&E, Ki-67, PCNA, and unknown dual-stained images from a variety of public websites. The results are very encouraging.

A study needs to be done to determine, say, for a given physical specimen, if, in fact, stain ratios can be automatically measured that are essentially independent of color-temperature or color balance used during capture, slice thickness, lamp brightness, camera gain, etc. This should be a blinded study. If the results are positive, this may prove to be a useful technique for reducing the variation in images taken at different times by different investigators for clinical studies, etc., allowing for more powerful studies with smaller sample sizes.

It should be noted that images to be studied should be flat-fielded, captured without compression, and captured without saturating the brightest values, if possible. Even quite dark values can be correctly assigned by the "V" algorithm.

---

references: (to be included in updated versions of this paper)

Correspondence should be addressed to "schuette@umich.edu"
or to the author at "MCIT, Arbor Lakes, Bldg 2, suite 1300,
4251 Plymouth Rd., Ann Arbor, MI, 48105, USA."