

Get started

Open in app



Follow

585K Followers



You have **2** free member-only stories left this month. [Sign up for Medium and get an extra one](#)

MICROSCOPY IMAGE PROCESSING ON WHOLE SLIDE IMAGES

Microscopy stain variations and how to estimate them

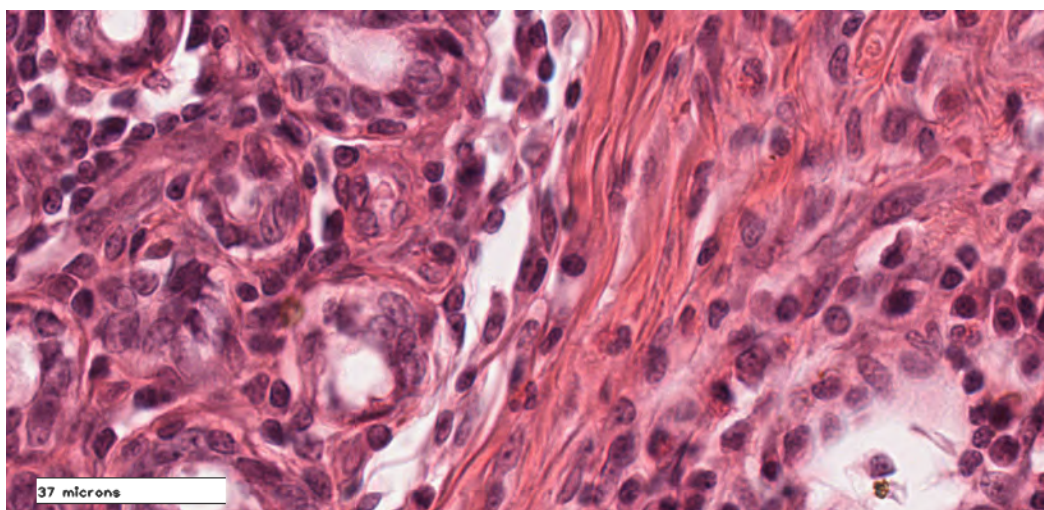
How Macenko's method works



Prof. Marc Aubreville · Jun 15, 2020 · 5 min read ★

If you have seen a couple of microscopy images of tissue (histopathology images), you must have noticed that they come in all variants of colors. Even when the same dying chemicals (stains) are used, the visual appearance is influenced by so many factors that it can easily become a big problem if you work with those images algorithmically.

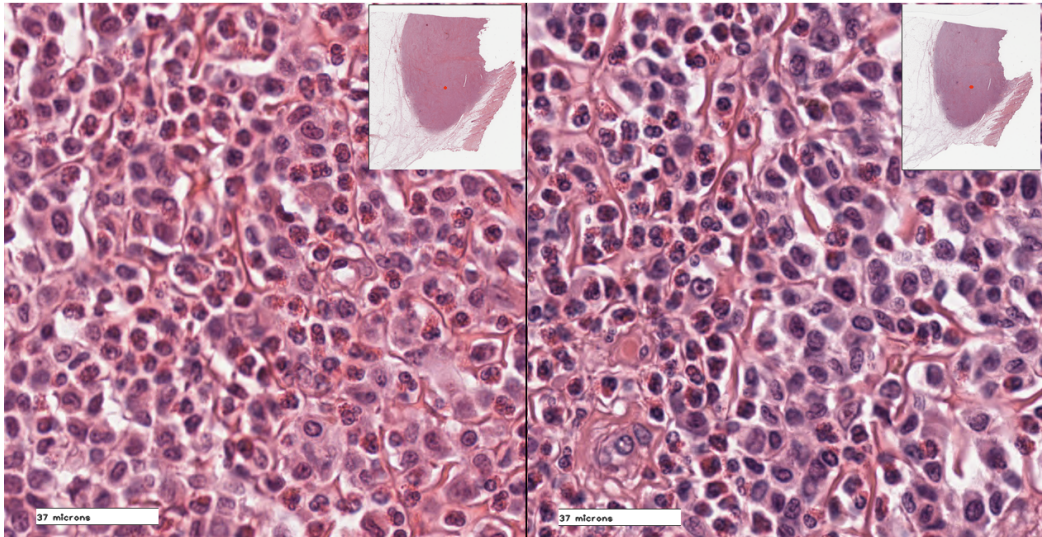
The default stain in histopathology is a combination of two chemicals: hematoxylin and eosin. The first is responsible for the dark violet (or blueish) color of all acid components (like, e.g., the DNA, residing in the cellular nucleus), while the second is responsible for the pinkish color that gives structure to the surrounding tissue. It looks for example like this:



A histopathology image of breast tissue. Thanks to C. Bertram and R. Klopfleisch from FU Berlin for providing the whole slide image. Image by author.

[Get started](#)[Open in app](#)

One major role in this game is played by the staining chemicals themselves. Depending on the concentration, the duration of staining and even the temperature. To stabilize these conditions, automatic staining machines are used. To showcase the influence of duration the tissue resides in a stain, have a look at the next image:



Histology image of canine cutaneous mast cell tumor, stained with H&E. Left image represents tissue left for 2 minutes in hematoxylin, while right side represents tissue left for 12 minutes in hematoxylin (Thanks to C. Bertram for providing the microscopy slides, image from author)

Both images represent slices of the exact same tumor, as can be seen in the overview thumbnail. However, the right image has a much stronger blue color component (caused by the much longer staining time in the hematoxylin solution).

If we want to process these images, these variations often pose a problem. Especially when we have only a small number of cases, it is at times hard to get our recognition algorithms robust, especially if they have a high pattern recognition capacity, such as by using deep networks.

To mitigate this, there are in principle two methods:

1. We can normalize the image before feeding it to the processing pipeline (stain normalization)
2. We can make the pattern recognition system more robust by artificially increasing the stain variability at its input (stain augmentation).

For both, it helps to be able to estimate the stain intensity of the two standard colors in the digital microscopy slide and modify it to either represent a standardized staining (in the case of normalization) or to represent many different stains.

Let us review the method of [Macenko et al.](#) [1] for this . While it might be a bit hard to understand at first, it is a very clever method that really deserves our attention.

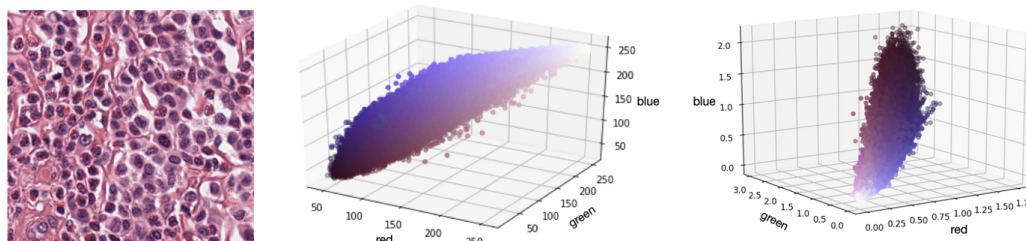
Get started

Open in app



was in the ray of light transversing to our CCD sensor, the darker the color. As you might know from radiology, light will also follow the [Beer-Lambert law](#) and be attenuated exponentially with the density of tissue. Macenko *et al.* thus convert each pixel in the image to a negative logarithmic representation (optical density, OD).

After this step, we have thus a linear superposition of both staining components — which is really cool for separating them! Also, higher numbers now represent more stain (and thus darker colors).



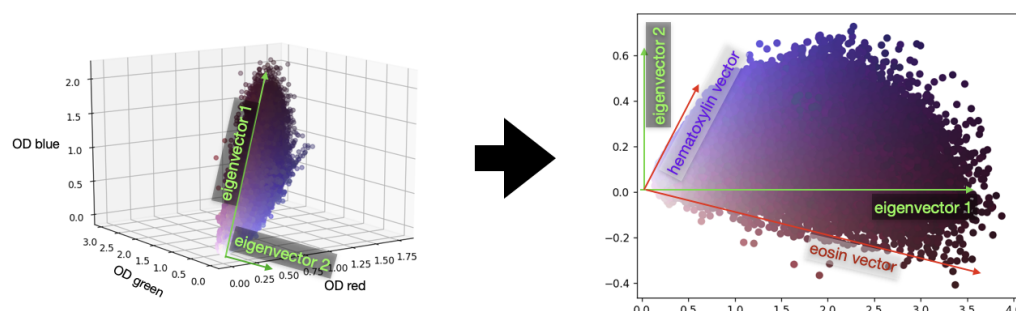
Original image, pixels in red, green and blue color space (middle) and in optical density color space (right).
(Image by author)

After removing too low values (which are only the bright background where no tissue ever was present), we can thus now try to estimate the staining components.

For this, they first estimate the 2D plane in the 3D space that is spanned by the two stain vectors. The singular value decomposition they use for this generates a new coordinate system, where the first component spans the main variance (which is, however, not yet one of the stain colors we are looking for). Each OD value of each pixel is being projected into the new coordinate system.

The beauty of this is that if we inherently removed noise by stripping away the third component (since we only used the first two eigenvectors). Why were we allowed to do so? Because we knew that the image is mainly just the colors by the hematoxylin and eosin (H&E) stain. Pretty cool.

Let us have a look at the same image, but only in 2D, for the sake of simplicity:



[Get started](#)[Open in app](#)


You may notice that I already added the two staining vectors in this image. They represent the maximum angle and the minimum angle in the coordinate system spanned by our eigenvectors. Why? Because they themselves span their own coordinate system — the actual H&E coordinate system.

In this image you can really nicely see that the color values are in fact a linear superposition of those two components. I already told you that we can estimate these (indicated by the red arrows in the image above) very easily now by calculating a robust minimum and maximum of the angle. You might argue that minimum and maximum are rarely robust, if we have some noise. That's a very good point and the reason why Macenko *et al.* used quantiles for this (typically the 1% and 99%).

But why is it exactly the minimum and maximum angles? Because in a noise-free world, you can not get lower than zero staining by the components, which is exactly what the area outside of these maximum angles would stand for.

In the next post, we will discuss how to apply this practically using Python — and how to do this on big images like the whole slide images we use in microscopy nowadays.

[1]: Macenko, Marc, et al. [A method for normalizing histology slides for quantitative analysis](#)” (2009) IEEE International Symposium on Biomedical Imaging: From Nano to Macro.

No rights reserved by the author. 

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

[Get this newsletter](#)

[Microscopy](#) [Data Science](#) [Pathology](#) [Deep Learning](#) [Microscopystainestimation](#)

[About](#) [Write](#) [Help](#) [Legal](#)

[Get the Medium app](#)

Get started

Open in app

