How Much Data is Enough?

For reasons of efficiency, it is important to understand how acquisition parameters impact file size. The major parameters influencing file size are 1) **bit-depth** (bytes/pixel), 2) **number of pixels per image**, and 3) **number of images** per time (e.g. frames/second) or space (e.g. frames/um in z). For example, a single channel, 1024x1024 pixel, 8-bit image requires 1024x1024 pixels x 1 byte/pixel = 1 megabyte (Mb) of storage, which can be captured and transferred to a computer in about 10 milliseconds. If time lapse imaging, 100 images can be captured in 1 second. The resulting file size is thus 1 megabyte/image x 100 images = 100 Mb, *in just one second of imaging*!

Beyond about 10 Gb (10,000 Mb), files become difficult to transfer and analyze. Transfer rates over Ethernet or to a magnetic disk are about 100 Mb/second, but can be substantially (5-10x) less. Thus, each minute of imaging time may translate into several minutes of data transfer time later. For analysis or rendering, standard desktop computers have about 10 Gb of free random-access memory available. Beyond 10 Gb, special hardware or software tools (that must be installed and learned) are required. Clearly, smaller files are better, but what is the impact on an image's scientifically relevant information content when the parameters that impact file size are reduced?

Bit-depth determines the number of light intensity levels that can be recorded at each pixel. For example, a sequence of 8 bits (0/1) can represent 2^8 = 256 intensity levels. If the goal is to count the number of photons per pixel, the bit-depth must be at least as large as the maximum number of photons detected. More commonly, the goal is merely to capture an image that looks nice on a computer monitor. In this case, the basis for comparison is the number of intensity levels that a computer monitor can display, which is about 256. Thus, when visual display is the main goal, 8-bits per pixel is more than sufficient to record all visually perceptible information.

The number of pixels in an image is related to image resolution. Reducing the number of pixels reduces resolution. While resolution is often a main goal, the presence of shot noise and aberrations inherently also reduces resolution, which can justify using fewer pixels. Reducing pixel resolution by 2x results in a 4x reduction in file size. Relatedly, if a small cell is sitting in the middle of an otherwise-empty coverslip, the field of view, and thus total pixels, can be reduced (cropped) to capture only the cell of interest.

Analogously, the number images acquired in z or t determine image resolution in depth and time, respectively. Again, the goal is to select a z-step or time lapse interval just big enough to capture the information that needs to be captured. If a cell only moves appreciably over one minute, there is no sense in acquiring an image every one second.

Finally, there are a wide-range of image compression algorithms that can further reduce file sizes after acquisition. Loss-less methods exactly preserve the acquired data, while lossy methods do not. Lossy compression should only be used with the help of an expert who understands how the compression will modify the image's scientifically relevant image content.

Collecting only the data that you need will save a fortune in the long run! Please contact us (microscopy@umich.edu) to help you decide how to most efficiently acquire the data that you need. We look forward to working with you!

Written by Dr. Aaron Taylor, BRCF Microscopy Core, December, 2019