Resolution and Sampling in Digital Imaging

Ian Dobbie

National University of Ireland, Galway, Ireland

ian.dobbie@nuigalway.ie

As nearly all imaging has moved from analogue media such as photographic film into digital image files it has become vital to understand how resolution and digital sampling affect the information contained in these images. This article is a basic introduction to the important concepts and how they should be applied, using general assumptions, rather than a rigorous mathematical approach. The focus is on optical microscopy although the essentials are the same for any digital imaging. The basic concepts are explained with reference to optical microscopy, but then the article moves on to more general aspects of sampling in digital imaging. More detailed, and mathematical, treatments can be found in a range of literature but a particularly detailed and instructive description can be found in Chapter 4, Handbook of Biological Confocal Microscopy, 3rd edition, Ed J.B. Pawley, Springer 2006. This is available as a sample chapter on the web at http://www.springer.com/0-387-25921-X.

Section 1: Resolution in a light microscope

The first section of this article covers resolution as produced in an optical microscope. It is important to understand this in order to appreciate the following section on sampling and digital imaging.

Images produced by a light microscope can be described in terms of two linked concepts: contrast and resolution. Contrast is the amount of difference in intensity, or brightness, between features. The human eye can tell contrast differences of about 10% between areas of different brightness. Resolution is a measure of how small an object can be and still be reliably visualised. This relies not only on the size of the objects, but also the contrast between them, hence there is a link between contrast and resolution. What's more, as features get smaller, the contrast they produce in the final image is reduced, further complicating the relationship between resolution and contrast.

The first thing to realise about a microscope, or any imaging system, is that it does not produce an exact enlarged copy of the sample under examination, but a representation of it. The microscope introduces distortions into the magnified image. Assuming that high quality optics are used, the most obvious of these distortions is limited resolution in the produced image. This limitation is imposed by diffraction and the best way to visualise this is to examine the Point Spread Function (PSF) of the imaging system. The PSF is the image produced by an infinitesimally small bright spot on a dark background. In a microscope this does not produce a single magnified spot in the image, but a bright disk with progressively weaker rings around it, an Airy Disk (Fig 1).

As a slight aside it should be noted that any structure can be said to be made up of a very large number of individual point like objects. So the image as a whole is made up of a sum of a number of Airy Disks with different intensities, but all with the same size, as their size is defined by the optics of the microscope.

The fundamental resolution of a microscope is then defined by how closely two objects, and hence two Airy Disks, can approach before they are indistinguishable from a single object (Fig 2A). As the two objects get closer to each other, the dip in intensity between their images gets progressively smaller, and they get harder to separate. The resolution of a microscope is conventionally defined by means of the Rayleigh Criterion. This states that two objects are no longer separable when the first minimum of the one Airy Disk falls on the peak of the second. In these circumstances, the dip in intensity between the peaks is 25% of the peak height (Fig 2B). The resolution using this measure is then the distance from the centre to the first minimum of the Airy Disk and is given by:

$$R = \frac{1.22\lambda}{NA_{cond} + NA_{obj}}$$

Where R is the resolution, λ is the wavelength of the light used and NA_{cond} and NA_{obj} are the Numerical Apertures (NAs) of the condenser and objective respectively. One important feature is that the resolution is independent of the magnification, but dependent upon the NA of the illumination optics (condenser), the NA of the collection optics (objective), and the wavelength of light used. A final important point related to the Rayleigh Criterion is that it is an arbitrary point at which to define a single "resolution" for a light microscope. There is information from smaller features within an image, but as features get smaller their contrast becomes progressively less and they become harder to image.

The fundamental reason for this limitation on resolution is diffraction. Images are formed in a microscope by recombining rays that enter the objective at different angles. Objects interact with the incident light and diffract it; the size of the object and the wavelength of the incident light determine by what angle the rays are diffracted. Smaller objects and longer wavelength light both lead to diffraction by a larger angle (Fig 3). This means that to detect smaller objects, rays diffracted at larger angles need to be collected, or a shorter wavelength light must be used.

In order to understand how the above equation relates to the actual process of collecting microscopic images it is necessary to know what the NA of a lens is. The NA is a measure of the range of angles over which light illuminates or is collected from the sample. The light coming from a lens forms a
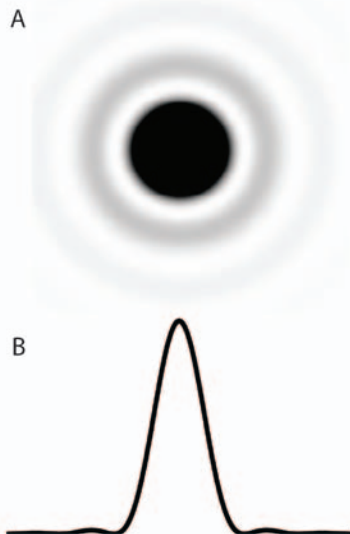

Fig 1. A) The pattern produced by a point like object from a microscope, an Airy disk. This image has had its contrast enhanced to emphasize the first trough and the subsequent rings around the central disk.

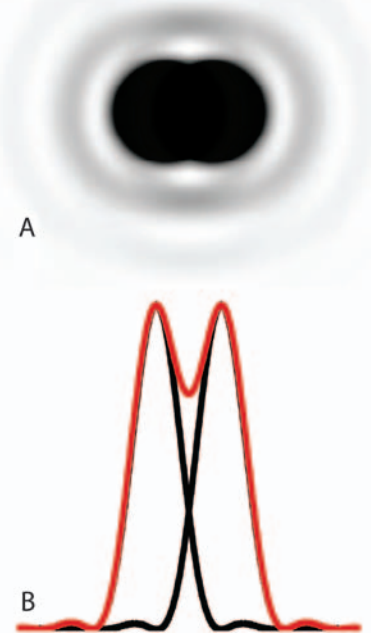*B) A cross-section slice through the centre of the Airy disk showing its full intensity profile.*

Fig 2. A) The image produced by two objects with a separation of the resolution as specified by the Rayleigh Criterion. The image is equivalent to two Airy disks, with the center of the second at the first minimum of intensity of the second.

B) A cross-section slice through the image in part A. The individual Airy disk traces are in black, with the total intensity shown in red.
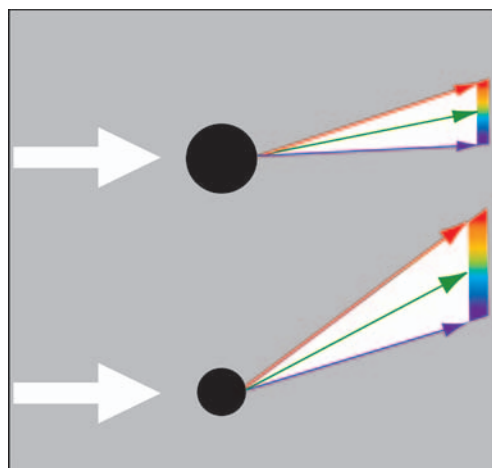
Fig 3. How objects diffract white light when they interact with it. A beam of white light, projected from the left, interacts with an object, which diffracts it into a spectrum of colours. Red light, with the longest wavelength is diffracted the most while violet light with the shortest wavelength is diffracted the least. A smaller object diffracts light by a larger angle.

cone, with the sample at the apex and the lens at the base of the cone. A low NA means a narrow cone of light is used, a larger NA means a wider cone is used (Fig 4A). The NA is defined as the sine of the half angle of the cone (in radians) times the refractive index of the immersion media (with air being very close to 1). As the NA is corrected for the immersion media used, it is a convenient measure to directly compare between lenses designed for different media such as an air objective and an oil immersion objective. Without a condenser, the maximum deflection of light that can be collected by the objective is defined by the objective NA. When a condenser is added to the microscope then the maximum possible deflection that can be collected is defined by the sum of the objective and condenser NA's.

Most resolution critical microscopy currently performed is performed using epi-fluorescence. This is a special case of the above discussion in that the objective is also used as the condenser and hence the denominator of the above equation becomes $2 \times NA_{obj}$.

As a first approximation for microscope imaging using a high NA lens, such as a Zeiss 63× 1.4NA PlanApo lens, the resolution is roughly $\lambda/2$. Where λ is the wavelength of the light used. A lens with a smaller NA will give lower resolution. This means that in the best case, with green (500 nm) light the resolution limit for an optical microscope is about 250 nm. The discussion so far focused on resolution in the *x-y* plane. In the z direction the resolution is also diffraction limited and is roughly 3× worse, in the best case this will be ~750 nm.
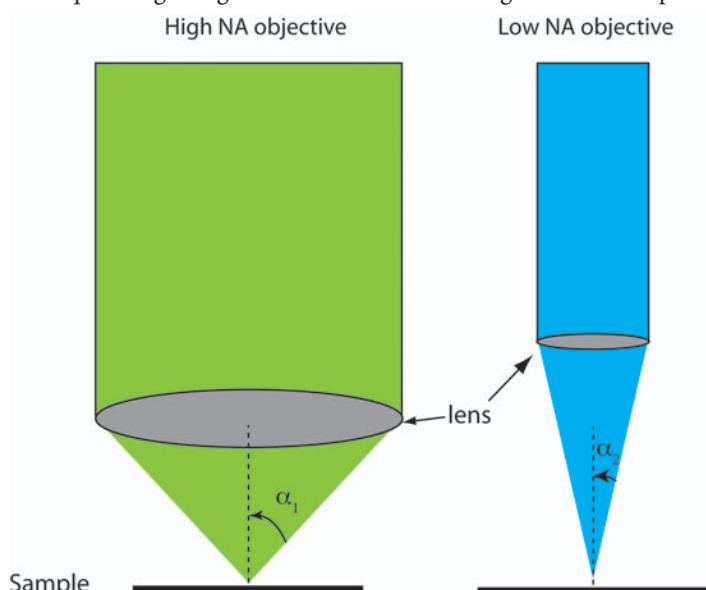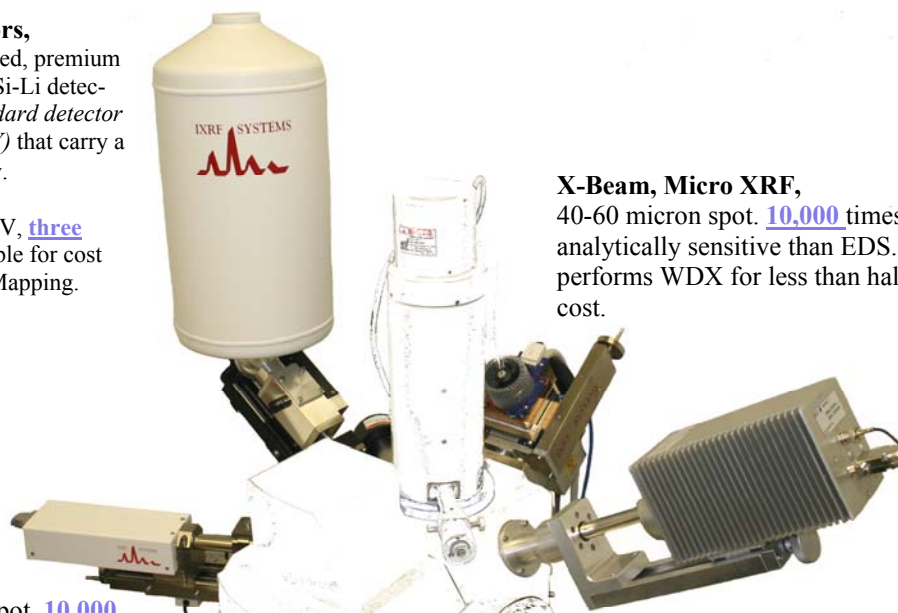
Section 2: Sampling and Digital imaging

The following section covers the more general ground of sampling an analogue signal into a digital image. Again, this discussion focuses on digital imaging in optical microscopy but the basic concepts are identical for any digital imaging application.

Applying the rules discussed previously, we have managed to generate a microscopic image with a resolution of 250 nm. How do we now sample this image in order to preserve its information content and not produce artefacts? The first thing to realise is that this resolution is as-measured in the sample. The image has been magnified by the objective, and so with a 63× objective the resolution in the primary image plane, where the detector is likely to be, will be 15.8 μm (250 nm × 63). A different magnification objective with the same NA will produce a different resolution at the primary image plane but have the same resolution at the sample. A lens with a different NA will produce a different resolution in the sample. If the detector is not at the primary imaging plane and a further magnifying lens is used then the resolution at the detector will be correspondingly larger.

From a naive point of view, sampling this with 15.8 μm pixels

would be enough to capture all the information contained. With reference to Figure 5A, in order to reliably separate these two peaks there must be at least 2 pixels across a resolution sized region. Further, consider that two objects, with their corresponding Airy Disks, are this distance apart but do not exactly line up with the pixel centres (Fig 5B). In this case there needs to be more than 2 pixels per resolution sized object. In order to reliably sample these two objects a detector with a pixel pitch of less than 8 μm/pixel is needed.

There is an extensive body of work related to this digitisation task from signal processing theory. This was developed in the early 20th Century because of its use in telecommunications. The most famous, and applicable piece of this work is referred to as Nyquist or Shannon/Nyquist theory. This states that a signal must be digitally sampled at more than 2 times the resolution (known as the bandwidth in signal processing), in order to be able to reliably reconstruct the original signal from its digital representation. Generally this is applied by sampling an image at 2.3× its resolution in order to preserve the content of the image. Signals digitised in this manner are said to be Nyquist Sampled.

There is a slight complication in that the resolution of a digital imaging system with square pixels is direction dependent. The spacing of the pixels is 1.41× √2 larger across the diagonal compared to that in the *x* or *y* directions. This suggests that the image needs to be sampled at more than 2.8× the resolution. Mostly this extra complication is ignored since the above discussion is based on the best possible case and in all realistic situations other factors will reduce the resolution. Calculating the necessary sampling for the example of a fluorescence microscope given above, with a 63× 1.4NA objective, leads to the requirement for a sensor with a pixel size of 6.9 μm.

What happens if the image is not sampled with pixels that do not properly match the resolution? If the detector pixel size is smaller than the resolution produced at the primary image plane, the image will be over-sampled. Conversely, if the pixel size is bigger, the image will be under-sampled. These two conditions have their advantages and disadvantages. Over-sampling means that there are more pixels than needed to encode the information in the image. This can lead to a nicer looking image, as features will be smoother. However, since the image is spread out over more pixels, there will be more detector noise. This can be especially bad at very low light intensities. Over-sampling can be useful for producing images that will be enlarged, but requires higher light intensities and can damage sensitive samples.

Fig 4. A) The Numerical Aperture of a high and a low NA lens. The half cone angles, α_1 and α_2 are shown for both lenses.
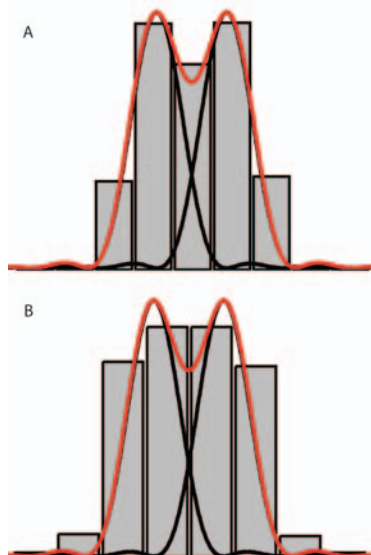
Fig 5. A) Digitisation of the image shown in Fig 2 with pixels at double the resolution. A slice through the image data, with total intensity shown in red, superimposed on the summed data as recorded by a digital image sensor. A) If the central dip in intensity falls exactly on a pixel, it is easily detected. B) If the central dip falls exactly between two pixels the dip is not detected at all. This shows that greater than 2x the image resolution is needed to separate these objects.

Under-sampling will mean that the high-resolution information is lost from the image, as features are spread over fewer pixels. The main advantage of under-sampling is in low light conditions, where the signal is concentrated on fewer pixels. This creates a signal that has a greater amplitude relative to background noise, and so boosts the signal to noise ratio. Under-sampling can also be advantageous in terms of image size versus field size. It is possible to sacrifice resolution for field of view, giving a large field without needing a huge number of pixels to cover it.

In general sampling will be determined by the camera pixel size and objective. Although these are not normally changeable (do you have several high specification cameras lying around?) it is often possible to increase the effective pixel size of the camera using binning. Binning is a mechanism allowing multiple adjacent pixels to be read out as one, reducing the effective number of pixels in the camera but increasing the pixel size. A square array of pixels, say 4, 9, or 16 pixels, are "binned" to produce a single pixel in the output image. This binning is done on the detector during image acquisition, meaning that readout noise is minimised.

A final point about digital sampling of images is an effect called aliasing. This occurs when imaging periodic structures that are close to or smaller than the resolution of the system. When aliasing occurs, wavy Moiré patterns can be superimposed on the periodic structure, or structures. The only real way to avoid this is to sample the image at higher resolution, *i.e.* with smaller pixels. Using properly calculated Nyquist sampling should avoid aliasing problems since the imaging system will effectively filter the image and eliminate the sub resolution information that creates the aliasing.

The above discussion on resolution and ideal pixel sizes have focused on using monochrome cameras to sample images. Adding colour to the mix complicates things greatly. Digital cameras are fundamentally sensitive to the incident light intensity, and not its colour. They do have a spectral response, which means that their sensitivity varies with wavelength (or colours) but they cannot tell blue from green. Most colour cameras achieve colour sensitivity by having a colour mask in front of the image sensor. This is an array of colour filters in front of the pixels. The most common pattern of filters used is the Bayer Mask (Fig 6). This has a 2×2 square pattern of 4 filters, one red, 2 green and one blue. The three colours, red, green and blue, match the colour sensors of the human eye and the 2×2 square makes it easy to tessellate over the image sensor. The green filter is used over 2 pixels in the mask as the human eye is significantly more sensitive to green than to either red or blue. This means that extra details in the green channel convey more information to the observer.

A camera using a Bayer Mask therefore has half the resolution in green and a quarter in both red and blue when compared to the raw image sensor. In order to produce a full resolution colour picture the extra data is then recalculated using interpolation between the pixels that are measured. This leads to a picture that looks as though it has the stated resolution but, in fact, there is significant information loss compared to the original image. This arrangement, although convenient, has significant disadvantages in terms of correctly producing a digital representation of the original object. In addition to creating an image at lower resolution than expected, there are two further problems. Firstly, these cameras are significantly less sensitive to light as the filters absorb at least 2/3rds of the incident light; secondly the data for the different colours is actually gathered from different parts of the image, which can lead to artefacts, especially with objects near the resolution limit.

The best way to gather a full resolution colour image is not to use a colour camera but to use a monochrome sensor and take three consecutive images with the relevant filters in front of the whole image sensor and then merge the three images to produce a full colour image. Using a camera with a Bayer mask, the image must be over-sampled by a factor of 2, in order to match the Nyquist sampling frequency for all three colours.
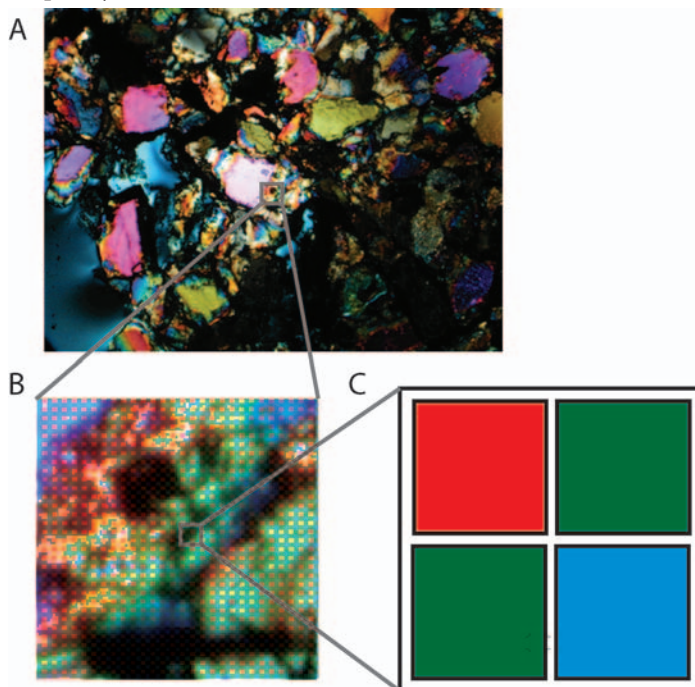

Fig 6. The Bayer mask. A) An image taken with a colour camera. B) How the colours match onto the Bayer mask. C) The individual colour pixels of one repeat at high magnification. Image kindly provided by A.G. Ryder, NUI-Galway, Ireland.

Conclusion

This article has discussed the pixel size at which digital images need to be sampled in order to observe diffraction-limited features. It is a complicated topic with many factors contributing to the final required digital sampling of the image. The basic rules necessary to optimise information gathering in digital imaging are given above, particularly focused on optical microscopy. In practise many factors other than resolution often contribute, such as low light intensities, meaning that experimentally it is necessary to sample with larger pixels and sacrifice resolution. As a final point it should be noted that although it is impossible to differentiate objects closer than the resolution limit, in the right circumstances nm level accuracy can be gained from optical microscopy (for example see Betzig *et al.* 2006, Science. vol 313, pg. 1642-5). ■