# **1** Digital Microscopy: Nature to Numbers

Ann Wheeler

Advanced Imaging Resource, MRC-IGMM, University of Edinburgh, UK

Bioimage analysis is the science of converting biomedical images into powerful data. As well as providing a visual representation of data in a study, images can be mined and used in themselves as an experimental resource. With careful sample preparation and precise control of the equipment used to capture images, it is possible to acquire reproducible data that can be used to quantitatively describe a biological system, for example through the analyses of relative protein or epitope expression (Figure 1.1). Using emerging methods this can be extrapolated out over hundreds and thousands of samples for high content image based screening or focused in, using emerging technologies, to data at the nanoscale. Fluorescence microscopy is used to specifically mark and discriminate individual molecular species such as proteins or different cellular, intracellular or tissue specific components. Through acquiring individual images capturing each tagged molecular species in separate channels it is possible to determine relative changes in the abundance, structure and - in live imaging - the kinetics of biological processes. In the example below (Figure 1.1), labelling of F-actin, a cytoskeletal protein, using a

#### Standard and Super-Resolution Bioimaging Data Analysis: A Primer, First Edition. Edited by Ann Wheeler and Ricardo Henriques.

<sup>© 2018</sup> John Wiley & Sons Ltd. Published 2018 by John Wiley & Sons Ltd.



Figure 1.1 Bioimage quantification to determine the dynamics of actin using photoconversion. Tsang, Wheeler and Wan Experimental Cell Research, vol. 318, no. 18, 01.11.2012, p. 2269–83.

fluorescent protein allows measurement of how fast it turns over in moving cells normally, and in a condition where a putative regulator of cell migration DSG3 is overexpressed. It shows that overexpressing DSG3 destabilises actin and causes it to turn over faster. Quantifying the expression and localisation of F-actin in several cells over time it is possible to see how much F-actin it turns over in the course of the experiment, where this happens, and the difference in rate between the two (Figure 1.1, graph). This type of scientific insight into the spatial and temporal properties of proteins is only possible using bioimage analysis and illustrates its use in current biomedical research applications.

In this book we are primarily going to consider quantification of images acquired from fluorescence microscopy methods. In fluorescence microscopy, images are acquired by sensors such as scientific cameras or photomultiplier tubes. These generate data as two-dimensional arrays comprising spatial information in the x and y domain (Figure 1.2); separate images are required for the z spatial domain – known as a z stack – which can then be overlaid to generate a 3D representative image of the data (Figure 1.2). Image analysis applications such as Imaris, Volocity, Bioimage XD and ImageJ can carry out visualisation, rendering and analysis tasks. The most sensitive detectors for fluorescence and bright-field microscopy record the intensity of the signal emitted by the sample, but no spectral information about the dye (Figure 1.3).



Figure 1.2 Workflow for bioimage data capture in 2D and 3D.



 
 Channel 1: Cell Junctions, E-Cadherin is the epitope labelled
 Channel 2: Microtubules, Alpha-tubulin is the epitope labelled
 Merged imaged - Channel 1 and Channel 2 combined together

**Figure 1.3** Combining channels in fluorescent bioimage analysis. Channel 1 has antibodies raised against E-cadherin labelled with AlexaFluor 568 secondary antibodies. Channel 2 is labelled with primary antibodies raised against Alpha tubulin and secondary antibodies labelled with AlexaFluor 488.

This means effectively that intensity information from only one labelled epitope is recorded. To collect information from a sample which is labelled with multiple fluorescent labels the contrast methods on the imaging platform itself - e.g. fluorescent emission filters, phase or DIC optics - are adjusted to generate images for each labelled epitope, all of which can then be merged (Figure 1.3). Some software will do this automatically for the end user. The final dimension that images can be composed of is time. Taken together, it is possible to see how a 3D multichannel dataset acquired over time can comprise tens of images. If these experiments are carried out over multiple spatial positions – e.g. through the analysis of multiwell plates or tilling of adjacent fields of view – the volume of data generated can considerably scale up, especially when experiments need to be done in replicates. Often the scientific question may well require perturbing several parameters, e.g. adjustment of different hypothesised parameters or structures involved in a known biological process. This means that similar image acquisition and analvsis needs to be used to analyse the differences in the biological system.



Figure 1.4 The Bioimage analysis workflow.

In these cases although setting up an automated analysis workflow makes sense, to manually quantify each individual image would take a considerable time and would require a substantial level of consistency and concentration. The programming of analysis pipelines does require some work initially but it can be seen as letting the computer automate a large volume of tasks, making the research process more reliable, robust and efficient. Indeed some applications now allow data processing in batches on remote servers, computer clusters or cloud computing.

Biomedical image analysis follows a given workflow: data acquisition, initialisation, measurement and interpretation (Figure 1.4) – which will be discussed in brief in this introductory chapter, followed by a more in-depth analysis in subsequent chapters.

# **1.1 ACQUISITION**

## 1.1.1 First Principles: How Can Images Be Quantified?

Before data can be analysed, it needs to be acquired. Image acquisition methods have been extensively reviewed elsewhere [1, 3, 4]. For quantification, the type and choice of detector which converts incident photons of light into a number matrix is important. Images can be quantified because they are digitised through a detector mounted onto the microscope or imaging device. These detectors can be CCD (charged coupled device), EMCCD (electron multiplying CCD) or sCMOS (scientific CMOS) cameras, or photomultiplier tubes (PMTs). Scientific cameras consist of a fixed array of pixels. Pixels are small silicon semiconductors which use the photoelectric effect to convert



Figure 1.5 How images are digitised.

the photons of light given off from a sample into electrons (Figure 1.5). Camera pixels are precision engineered to yield a finite number of electrons per photon of light. They have a known size and sensitivity, and the camera will have a fixed array of pixels. Photons of light pass from the object to become images through the optical system, until they collide with one part of the doped silicon semiconductor chip or pixel in the camera. This converts the photons of light into electrons which are then counted. The count of 'photo electrons' is then converted into an intensity score, which is communicated to the imaging system's computer and is displayed as an image (Figure 1.5). PMTs operate on similar principles to scientific cameras, but they have an increased sensitivity, allowing for the collection of weaker signals. For this reason they are preferentially mounted on confocal microscopes. Photomultipliers channel photons to a photocathode that releases electrons upon photon impact. These electrons are multiplied by electrodes called metal channel dynodes. At the end of the dynode chain is an anode (collection electrode) which reports the photoelectron flux generated by the photocathode. However, the PMT collects what is effectively only one pixel of data, therefore light from the sample needs to be scanned, using mirrors, onto the PMT to allow a sample area larger than one pixel to be acquired. PMTs have the advantage that they are highly sensitive and, within a certain range, pixel size can be controlled, as the electron flow from the anode can be spatially adjusted; this is useful as the pixel size can be matched to the exact magnification of the system, allowing optimal resolution. PMTs have the disadvantage that acquiring the spatial (x, y and z) coordinates of the sample takes time as it needs to be scanned one pixel at a time. This is particularly disadvantageous in imaging of live samples, since the biological process to be recorded may have occurred by the time the sample has been scanned. Therefore live imaging systems are generally fitted with scientific cameras and systems requiring sensitivity for low light and precision for fixed samples often have PMTs. (https://micro.magnet.fsu.edu/primer/ digitalimaging/concepts/photomultipliers.html)

#### 6 STANDARD AND SUPER-RESOLUTION BIOIMAGING DATA ANALYSIS

## 1.1.2 Representing Images as a Numerical Matrix Using a Scientific Camera

Although having a pixel array is useful for defining the shape of an object it doesn't define the shading or texture of the object captured on the camera. Cameras use greyscales to determine this. Each pixel has a property defined as 'full well capacity'. This defines how many electrons (originated by photons) an individual pixel can hold. An analogy of this would be having the camera as an array of buckets, which are filled by light. It is only possible to collect as much light as the pixel 'well' (bucket) can hold; this limit is known as saturation point. There can also be too little light for the pixel to respond to the signal, and this is defined as under-exposure.

The camera can read off how 'full' the pixel is by a predetermined number. This is defined as the greyscale. The simplest greyscale would be 1-bit, i.e. 0 or 1. This means that there is either light hitting the pixel or not; however, this is too coarse a measure for bioimage analysis. Pixels record intensity using binary signals, but these are scaled up. Pixels in many devices are delineated into 256 levels, which corresponds to 2<sup>8</sup>, which is referred to as 8-bit. The cone of a human eye can only detect around 170–200 light intensities. So a camera, set at 8-bit (detecting 256 levels) produces more information than an eye can compute. Therefore, if images are being taken for visualisation, and not for quantification, then using a camera at 8-bit level is more than adequate. For some basic measurements, 8-bit images are also sufficient (Figure 1.6).



**Figure 1.6** Basic quantification of cellular features using 8-bit fluorescent image of F-actin.

It is possible to increase the sensitivity of the pixel further, currently to 12 (4096 or 2<sup>12</sup>), 14 (16384 or 2<sup>14</sup>) and 16 (65536 or 2<sup>16</sup>) grey levels. For detecting subtle differences in shading in a complex sample, the more numerical information and depth of information that can be mined from an image the better the data that can be extracted can be. This also allows better segmentation between noise inherent in the system and signal from the structure of interest (Figure 1.6).

Although this chapter is concerned with bioimage analysis it is essential that the images are acquired at sufficient sensitivity for quantification. Scientific cameras currently can delineate up to 2<sup>16</sup> grey levels dependent on their specification. The image histogram, is a 1D representation of the pixel intensities detected by the camera. It can be used to determine the distribution of pixel intensities in an image, making it easy to perceive the saturation or under-sampling of an image acquired (Figure 1.7). A saturated signal is when the light intensity is brighter than the pixel can detect and the signal is constantly at the maximum level. This means that differences in the sample can't be detected as they are being recorded at an identical greyscale value, the maximum intensity possible (Figure 1.7). Under-sampling, which means not making use of the full dynamic range of the detector or having information below the detection limit of the detector is not ideal. It means that the intensity information is 'bunched together', and so subtle structures may not be able to be detected (Figure 1.7). Under-sampling is sometimes necessary in bioimaging, for



- Data is squashed
- Can't tell differences between staining

 More sensitivity for specific measurements

Figure 1.7 The effect of saturation and under-sampling on bioimage analysis.

instance if imaging a very fast process or when a very weak signal is being collected from a probe which can be photo-damaged. Provided that sufficient signal can be collected for quantitative analysis this need not be a problem. However, best practice is to have the signal fill the whole dynamic range of the detector.

The first and perhaps most important step in bioimage analysis is that images be acquired and quantified in a reproducible manner. This means:

- using the same piece of equipment, or pieces of equipment that are technically identical
- ensuring equipment is clean
- ensuring samples are as similar as possible and prepared similarly
- using the same parameters to acquire data, e.g. same magnification, same fluorescent labels and very similar sample preparation and mounting.

# 1.1.3 Controlling Pixel Size in Cameras

Pixels in scientific cameras are a predefined size, while in PMTs the scan area can be adjusted so that pixel size can be varied (see Section 1.1 on acquisition). The ideal pixel size matches the Nyquist criteria – that is, half the size of the resolution that the objective permits, providing the pixel is sufficiently sensitive to detect the signal of interest. Camera pixel size can limit resolution as it is difficult to spatially separate two small structures falling in the same pixel unless subpixel localisation methods are used, as discussed in Chapter 8. It is very difficult to spatially separate two small structures falling in the same pixel. If a larger pixel size is required it is possible to have the detector electronically merge pixels together. This is generally done when a  $2 \times 2$  array of pixels or a  $4 \times 4$ array is combined into one super-pixel. The advantage of this is that there is a 4  $(2 \times 2 \text{ bin})$  or 16  $(4 \times 4)$  fold increase in sensitivity since the 'merged pixels' add together their signals. The trade-off is a loss of spatial sampling as the pixels are merged in space. For studies of morphology, the resolution of the camera is important; pixels (i.e. the units comprising the detection array on the scientific camera) are square, and for any curved phenomena the finer the array acquiring it, the better will be the representation curves of the sample. The loss of spatial detail can be problematic if the structures studied are fine (Figure 1.8). Using brighter dyes - that is those with a higher quantum yield of emitted photons per excited photon - and antifade agents to prevent bleaching can help here.



Figure 1.8 Binning of pixels to increase speed and sensitivity of Bioimage acquisition.

For studies of protein expression, sensitivity can be important, although the bit depth of the pixel plays a role. If the detector can only detect a fraction of the light being produced because it either meets its saturation point or is under-exposed it causes issues. The epitope will be either not detected or under-sampled because the detector is not capable of picking up sufficient signal for quantification (Figure 1.8).

In studies of fast transient reaction (e.g. calcium signalling), fast exposure and frame rate can be more important than spatial resolution (Figure 1.8). Here, binning can be extremely useful since the sensitivity to an individual pixel may not be sufficient to detect subtle changes in signal. Binning also allows the camera to record data and transfer this electronic information to the computer faster since there are fewer pixels (Figure 1.9).

Detectors have a finite capacity for signal and a certain output speed, and this can be analogised to an array of buckets that have a certain capacity for water and tip it out at a certain rate (Figure 1.10). Knowing the speed of the camera to write the detected information to the computer's disk is important. In live experiments, cameras can detect signals faster than the speed with which the computer can write information to the disk. This is known as a clocking problem and is troublesome because data is collected, but it isn't recorded to the computer disk (Figure 1.9).



Figure 1.9 Bucket brigade CCD analogy (Courtesy of Molecular Expressions, Florida state University, USA, https://micro.magnet.fsu.edu/primer/index.html).

56	76	64	78	34
72	63	89	76	97
82	47	102	83	83
68	83	79	72	65
54	66	52	64	55
Madia	- 70			

Median = 79

Figure 1.10 A  $3 \times 3$  median filter kernel. The filter size is indicated in orange. This filter smooths the image and denoises it.

The most recent advance in camera technology, sCMOS cameras, can be beneficial because they combine a small pixel size with high sensitivity and fast read time (clocking). They have applications in a wide variety of biological questions where the phenomena to be imaged are small and either transient or entail rapid kinetics. These devices can also be implemented for scanning of large areas in techniques such as light-sheet microscopy due to their large field of view and high-speed acquisition.

Camera manufacturers producing instruments that are suitable for quantitative imaging:

- 1. Andor Technologies http://www.andor.com/
- 2. Hammamatsu http://www.hamamatsu.com/
- 3. Leica Microsystems http://www.leica-microsystems.com/home/
- 4. Lumenara https://www.lumenera.com/
- 5. Nikon Instruments https://www.nikoninstruments.com/
- 6. Olympus http://www.olympus-lifescience.com/en/
- 7. PCO Instruments https://www.pco-tech.com/
- 8. Photometrics http://www.photometrics.com/
- 9. QImaging http://www.qimaging.com/
- 10. Motic Instruments http://www.motic.com/As\_Microsope\_cameras/
- 11. Zeiss Microscopy http://www.zeiss.com/microscopy/en\_de/software-cameras.html

# **1.2 INITIALISATION**

Initialisation is the step where bioimages are prepared for quantification. In most cases, the image generated by the system will not be immediately suitable for automatic quantification, and most analysis requires the computer to have a set of very similar artefact-free images for the analysis algorithms to function correctly. It is thus critical to minimise image features that may corrupt or hamper the analysis framework to be used. The dominant aberrations in the detection system are caused at three levels: (a) the sample itself, (b) the microscope or scanner's optical properties through which the image is formed and (c) the detector. These aberrations need to either be minimised or removed entirely so that the signal to be processed in the image is clearly distinguished from the noise which is otherwise present in the sample. Techniques used to do this such as filtering, deconvolution and background subtraction, and registration in x, y, z and colour channels needs to be carried out.

# 1.2.1 The Sample

The sample to be imaged may contain artefacts or structures that are challenging to image, which makes it difficult to acquire good images for analysis. The key to good analysis is excellent sample preparation. Dyes and antibodies need to be optimised so that they are bright enough to be within the linear range of the detector. Ideally the background from non-specific binding or antibodies or other probes would be reduced. The fixation and processing of samples would be optimised. Even with these strategies in place, a digital camera can only acquire a 2D image of a biological structure which is itself 3D. This means that out of focus light from around the focal plane is present in the image, which may obscure the signal from in-focus light. Confocal systems minimise out-of-focus light in acquired images by physical methods involving the use of pinholes. However, since most light in a sample is out of focus, only a small fraction of light is allowed through the pinhole increases the need for bright labelling [1]. Further inappropriate fixation or storage can damage samples, and sample mounting is also challenging because 3D samples can be squashed or shrunk. For studies in thick tissue, where the sample will be cut into a sequence of individual thin slices that will be imaged, there can be issues with collating these images back into a virtual 3D representation of the tissue [2].

# 1.2.2 Pre-Processing

Not all parts of images may need to be processed, and the regions to be measured may need to be turned into separate images. The imaging system may acquire data in a format that is not compatible with the analysis algorithm. Some imaging applications store images in individual folders (Leica LAS, Micromanager) and data may need to be moved to an analysis server. Due to the nature of image acquisition rescaling, techniques such as histogram equalisation may be necessary. All of these steps contribute to the pre-processing. Most applications enable this and would have some kind of image duplication function or a means of saving the pre-processed data separately from the raw data. The raw image data must be retained to comply with scientific quality assurance procedures which are discussed in Chapter 10, which deals with presentation and documentation.

# 1.2.3 Denoising

Denoising is removal or reduction of noise inherent in the sample and imaging system which masks the signal of interest. Cameras and PMTs are not perfect, and are subject to several sources of noise. Noise is defined as electrons that are read by the camera that have not been generated by photons from a sample, for example,

- *Shot noise:* This is caised by random electrons generated by vibration inside the camera or PMT.
- *Dark current:* PMTs and cameras have a baseline number of electrons that it reads even when there is no light. Manufacturers will usually set this to be a non-zero value, and PMTs in particular have a base current from photocathode to anode even in the absence of light. Measuring the dark current on a system is useful, because if this value falls below the normal value, it helps the end user determine that there is a problem with the camera. A low dark current can be achieved by cooling the detector; often CCD and EMCCD cameras are cooled for this reason.
- *Read noise:* The photoelectric silicon semiconductor has a range of accuracy, e.g. although it will usually generate two electrons per photon sometimes it may generate one and sometimes three. The accuracy of the read noise depends on the quality of the pixel chip. The number of electrons yielded per photon can be described as the quantum yield.
- *Spectral effects:* Neither PMTs nor cameras produce a linear number of photoelectrons per incident photon across the visible spectrum. At 500 nm, a camera may produce four electrons per photon and at 600 nm it may produce three and at 700 nm, just one. If correlations are being made between two different dyes or fluorophores, it is important to take into consideration what the 'spectral performance' of the detector is.
- *Fixed pattern noise:* Some cameras have random noise caused by spurious changes in charge across the pixel array. Other types, sCMOS in particular, suffer from fixed patter noise, which means that, due to manufacturing or properties of the camera itself, certain parts of the camera have a higher noise level than others. This is often in a fixed pattern, although it can consist of individual 'hot' (i.e very noisy) pixels. This noise pattern can be subtracted from an image.

All scientific cameras and PMTs from reputable manufacturers will include a table and datasheet describing the performance of their instruments. This can be useful to study at the outset of an experimental series where Bioimage analysis is to be done.

## 1.2.4 Filtering Images

Noise is inherent in all bioimages; this may be introduced because of shortcomings with the detector as described above. This type of noise is described as non-structural background, and is low-frequency, and constant in all images. Another source of noise is introduced because the detector can only acquire images in 2D while biological samples are 3D, so out-of-focus light, or issues with labelling the sample may cause the desired signal to be masked. This type of noise is high frequency and can have structural elements. One of the most frequently used methods for initialising images for bioimage analysis is filtering. By using a series of filters it becomes possible to remove most of the noise and background, improving the signal-to-noise ratio. This is generally achieved by mathematical operations called deconvolutions.

In a nutshell, this involves deconvolving the numerical matrix that makes up the bioimage with another number array; they can contain different numbers depending on the desired effect on these images. The technical term for these arrays is kernels, and denoising involves filtering images using kernels.

Detector noise and non-homogenous background from the sample can be removed by a process called flat fielding. This is acquiring an image with a blank slide at the settings used to acquire the bioimages, and subtracting this background noise image from the data. Some image analysis programs can generate a pseudo flat field image if one has not been acquired. This method can be very effective with low signal data if the noise is caused by the detector. 'Salt and pepper' noise can be evened out by using a median filter. A median filter runs through each pixel's signal, replacing the original pixel signal value entry with the median of its neighbours. The pattern of neighbours is called the "window" (Figure 1.10).

The effect is nonlinear smoothing of the signal, but edges of the images suffer as the median value of the edge will involve a null value, which means that a few edge pixels are sacrificed when using this method. Often images generated from PMTs suffer from this type of noise because of shot noise and read noise on the detectors. Other types of filters that can reduce noise in samples are as shown in Figure 1.11a:

- Smooth filter: A pixel is replaced with the average of itself and its neighbours within the specified radius. This is also known as a mean or blurring filter.
- Sigma filter: The filter smooths an image by taking an average over the neighbouring pixels, within a range defined by the standard deviation of the pixel values within the neighbourhood of the kernel.

(a)









Original

Sigma filter

Gaussian blur

Median filter

(b)

(c)



Before correction

After correction



Before correction ROI in orange

After correction

**Figure 1.11** Initialisation using filtering (a) Illustrative example of image filtering taken from the Image J webpage https://www.fiji.sc, (b) Example of rolling ball background subtraction: left-hand side is before correction, and right-hand side after, (c) Using ROI subtraction.

• Gaussian filter: This is similar to the smoothing filter but it replaces the pixel value with a value proportional to a normal distribution of its neighbours. This is a commonly used mathematical representation of the effect of the microscope on a point of light.

In epifluorescence images there is often a vignette of intensity across the image. This is a result of the illumination in these systems where a mercury halide or LED illuminator is focused into the centre of the field of view to be imaged, provided it is correctly aligned. The bulb will not give an even intensity of illumination; rather the illumination follows a Gaussian distribution. In well-aligned microscopes this means that the image is brightest in the centre and dimmer at the edges. If there is a problem with the alignment of the illuminator, there can be an intensity cast across the image where potentially one of the corners or part of the image is brighter than another. To remove this issue, in Image J a 'rolling ball' background correction algorithm designed by Castle and Keller (Mental Health Research Institute, University of Michigan) is implemented (Figure 1.11b). Here a local background value is determined for every pixel by averaging over a very large kernel around the pixel. This value is hereafter subtracted from the original image, hopefully removing large spatial variations of the background intensities. The radius should be set to at least the size of the largest object that is not part of the background [3].

In better-aligned systems or systems which inherently have more even illumination such as confocals, noisy background can be caused by other effects. For instance, uneven illumination caused by "scan lines" in confocal transmitted light images can be removed using the native FFT bandpass function present in ImageJ and other software packages. When detector noise or bleaching is an issue, this can be accounted for by measuring the mean intensity of the region in an image where there is known background and then subtracting the mean value of this region. Although this reduces the net intensity value in an image, it can emphasise relevant data (Figure 1.11c). Removing the high frequency noise caused by labelling, light interference in the sample can be more challenging. Different types of filters can assist with this, and this subject is discussed at greater length in Chapter 3.

## 1.2.5 Deconvolution

Deconvolution is a method which is used to remove out-of-focus light completely from an image. It is based on the premise that an image is a convolution of the imaged sample with the system used to image it – in the case of light microscopy, the sample and the microscope. No system is optically perfect and objective lenses are a primary cause of aberrations in an image. They suffer from multiple aberrations, predominantly spherical and chromatic, and have artefacts in flatness of field. Highquality objectives such as Plan-Apochromat are corrected for all of these across the visible spectrum but are more expensive than most other objectives. In particular, aberrations in the axial dimension can be particularly problematic for light microscopes. Any lens may do a fairly reasonable job of focusing light in 2D, but 3D focus is more challenging, and lenses tend on average to perform half as accurately in the third (axial) dimension as in x and y. Abbe's law summarises the resolution of a light microscopy image.

$$d = \frac{\lambda}{2n\sin\theta}$$

where *d* is the resolution of the system,  $\lambda$  is the wavelength of emitted light, *n* sin $\theta$  is the numeric aperture of the objective (the numeric aperture is the half angle that light can propagate through the objective). Only the most simplistic of imaging system consists of just an objective. Fluorescent systems will also have dichroic mirrors and filters as well as other moving parts. All of these will slightly distort or absorb photons on their path to the detector. Each distortion, though incremental, adds up. Experts in optics tend to combine this source of error in the optical system in one metric: the point spread function (PSF). This describes how an optical system images an infinitely small and perfectly spherical point of light (Figure 1.12). Inevitably, the refraction of light due to imperfection in the optical system will mean that the point of light is distorted.

Since all images consist of many points of light, knowing about how a given imaging system distorts one point of light means that it's possible to extrapolate this onto an image and 'deconvolve' out the real signal from the distortions introduced by the imaging system [4]. Most bioimage processing applications provide some type of deconvolution, although specialist packages such as Autoquant or Huygens specialise in this. Many depend on an artificial point spread function which will be generated based on the wavelength of emitted light and the magnification and numerical aperture of the objective used. Naturally, contributions from dirty lenses or misaligned optics are not taken into account when generating an artificial PSF. For heavily used instruments



**Figure 1.12** Experimental point spread functions: By Howard Vindin (own work) [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons.

it may be advisable to generate an imaged PSF, for example through the observation of small-scale fluorescent beads (e.g. 100nm Tetraspeck beads – Thermo Scientific). Several algorithms are provided for deconvolution, the most commonly used being:

- Nearest neighbour approaches: These are fundamentally two-dimensional, they are classified for the purposes of this discussion as deblurring algorithms. As a class, these algorithms apply an operation plane by plane to each two-dimensional plane of a three-dimensional image stack. For example, the nearest-neighbour algorithm operates on the plane z by blurring the neighbouring planes (z+1 and z-1, using a digital blurring filter), then subtracting the blurred planes from the z plane. This has the disadvantage that it can reduce the signal-to-noise ratio, and it can add in sharpened points caused by overlapping signal in the z domain in places where it doesn't belong. The advantage of this approach is that it's computationally light, at the expense of degrading the signal and introduction of artefacts.
- *Iterative deconvolution:* This is an image restoration approach where out-of-focus light is either discarded or brought into focus. The algorithm works by iterating through a set of parameters that best represent the 'in focus' image. To start with, an estimate of the object is performed; generally this is the raw image. This estimate is convolved

with the PSF, and the resulting 'blurred estimate' is compared to the original input image. A metric (figure of merit) of restoration indicates the improvement in the images compared to the original. This metric is then used to adjust the estimated parameters to reduce the error criterion in the subsequent iteration. This process is repeated until the error criterion is minimised or reaches a specified threshold. The final image is the object estimate at the last iteration. Richardson–Lucy deconvolution is a popular implementation of this method [5, 6]. These iterative algorithms often use the likelihood error criteria defined in probability theory. Maximum likelihood estimation (MLE) is a method of *estimating* the parameters of a statistical model given certain observations, e.g. the blurred object image, by finding parameters that maximise the likelihood of making the observations given the parameters. Several commercial applications use these approaches: Huygens, Imaris, Nikon Elements, Carl Zeiss and ImageProPremier [7].

Iterative statistical algorithms are more computationally intensive when compared to non-iterative methods and can take significantly longer to reach a solution. However, they may restore images to a slightly higher degree of resolution than filtering. These algorithms also have the advantage that they impose constraints on the expected noise statistic (in effect, a Poisson or a Gaussian distribution). As a result, statistical algorithms have a more subtle noise policy than simply regularisation, and they may produce better results on noisy images. However, the choice of an appropriate noise statistic may depend on the imaging condition, and some commercial software packages are more flexible than others in this regard. Processing on a server for deconvolution can be a major advantage for a large batch of deconvolution datasets, as they will often take up a significant amount of processing capability of a desktop machine. Further information about deconvolution can be found here:

https://micro.magnet.fsu.edu/primer/digitalimaging/deconvolution/ deconvolutionhome.html

#### 1.2.6 Registration and Calibration

Spatial calibration of images: Pixels all have a defined size so the amount of physical space of images on a detector will depend on the optical elements (e.g. the microscope lens) that are projecting light onto it. If a small ruler, called a graticule, is imaged it is possible to determine, for a given optical system (e.g. a tissue culture microscope with  $10 \times$  lens), what the actual size detected by each pixel is, that is, if  $100 \mu m$  on the graticule is visualised across 50 pixels then one pixel will measure across  $2 \mu m$ . Many commercial systems will come pre-calibrated, and this information will be included in the image header file, but home-built systems may not be pre-calibrated, and systems where the detector or objective lens has been changed for experimental purposes may also not be correctly spatially calibrated. If you have a high-precision stage, it becomes possible to calculate the pixel size calibration by moving an object for a well-defined distance in the stage. The observed distance travelled in images can then be correlated to the distance set on the stage to get the pixel calibration. For measurement of spatial parameters such as size and shape of imaged objects it is essential to include these measures.

Image registration is technically described process of 'overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different instrument settings (e.g. different fluorescent channels)' [8]. Registration geometrically aligns two or more images, which can comprise different channels or different x and y or z planes. Registration algorithms may need to be applied in the following cases:

- Datasets where more than one fluorescent channel is used. Chromatic aberrations will mean that the colours may focus to slightly different places on a camera or detector.
- Datasets collected over time, where the structure may move or the system may drift.
- Datasets collected by multiview analysis, e.g. tiling or mosaicing of an image.

Each of these cases is very different, and the algorithms programmed to address these differ; however, they do work on similar workflows whereby features are detected and matched, the geometric functions required to map or transform the image are created, and the geometric transform is applied. Registration of many biological images can be challenging due to the structures involved and the need to align spatial positions and different channels; image processing experts have generated several tools for this [9]. Libraries, plugins and other tools for image registration and stitching are supplied in software packages such as Amira, Arivis and Vaa3D for this. In some cases the software may ask for features or fiducial markers which can be used to remap the images [10]. In other cases autocorrelation routines or propagation methods can be used, particularly to address system drift. Several opensource plugins for registration are available and can be installed into ImageJ/Fiji (http://imagej.net/Category:Registration). Often these will have been developed with specific use cases in mind: e.g. subpixel registration of super-resolution microscopy images; stitching and multiview reconstruction of large tissue or cell areas; registration of light sheet microscopy data. Some examples of registration algorithms are: TrakEM2 [11], SURF+affine transformation [12] UnwarpJ, [13] and V3D, [14] and BrainAligner [15].

#### **1.3 MEASUREMENT**

Once the noise has been removed from the image, generating useful numerical data for measuring samples can be started. The pixel array from an imaging device, CCD, EMCCD or sCMOS camera or PMT generates a numerical matrix. Hence bioimage measurement is, simply put, sampling this matrix and performing mathematical operations yielding numeric descriptors of the data. Once the image is corrected for any inherent aberrations and is pre-processed so that it is as close a representation of the original object imaged as possible, it is time to move forwards and gather numerical data.

For these features to be measured they must be segmented (i.e. identified as interesting areas to analyse, e.g. nuclei are segmented for imaging when nuclei counting occurs) and then quantified. This topic is very complex, which is why Chapters 2 and 3 of this book are dedicated to the subject, but it lies at the heart of bioimage analysis. Typical features in a light microscopy image to be measured would be: Size and shape of biological structures, as a non-exhaustive list: nuclei, endosomes, cytoskeletal components in cells. In tissues it might be neurons, blood vessels or populations of stem cells. Location where certain epitopes are, in respect to reference organelles, is often asked. One of the simpler use cases is whether a given protein is localised in the nucleus or in the cytoplasm, although this can be extrapolated to many cases, in particular in developing organisms. Motion or kinetics of structures could be cells crawling, delivery of cargo on a microtubule or looking at cells developing their fate. Concentration, the amount of a given epitope in a specific location in a cell or tissue, is a frequently asked research question. This can be used for examples applied to fluorescence recovery, after photobleaching (FRAP), which would look at how much protein turns over in a given place. Concentration studies are also useful when investigating the role of individual components of a biological complex to its function, e.g. if protein x is removed does the complex still form or only partially form?

#### 22 STANDARD AND SUPER-RESOLUTION BIOIMAGING DATA ANALYSIS

These parameters in image analysis would generate the following types of numerical output:

Measuring shape or location would generate Cartesian coordinates (x,y,z) and metrics describing area, length width or the perimeter of a structure, e.g. a nucleus (Figure 1.13). This is done by selecting a tool which puts a contour over the image and the number of pixels it intersects with – or for area are inside a closed contour – are counted. With area the number of pixels intersected by the line and the number of pixels inside this area are also counted.



**Figure 1.13** Using ImageJ to select parameters of shape and intensity in an image of nuclei (blue). Here nuclei have been manually segmented using a contour – yellow line. Measurements have been set in ImageJ, and the numerical results output. The area of each nucleus, mean, standard deviation, maximal and minimal intensity are computed. The circularity (Circ) of the nuclei are also computed.

Measuring intensity: Here instead of the number of pixels being counted, the intensity value of each of the pixels is counted. The mean, maximum, minimum and standard deviation of these values can easily be determined. It is also possible to sum the values up to give an integrated sum for this (Figure 1.13).

**Measuring kinetics:** This is a convolution of measurement of either shape or intensity with the time domain.

There is a very wide variety of both commercial and open-source software which can perform measurements on bioimages. A list which covers most of the frequently used applications is given in Table 1.1.

Once the measurement parameters have been developed they can be stored and applied to large data samples. Specific software has been written to better support these high content or big data applications; these are discussed in the Chapter 9.

#### **1.4 INTERPRETATION**

Once the most interesting and relevant metrics for a given question have been identified statistical analysis can be applied to this. This can be done through packages such as SPSS, R, Graphpad, Prism, MATLAB, Minitab or by using bespoke analyses. There will be different statistical methods for analysing the spatial distribution of clusters or determining whether one or more features in an image correlates or co-locates with another, and the relevant tests will depend on the question being asked. These types of analysis are described in Chapter 6 and 8. Bioimage quantification analysis has historically relied overmuch on the t-tests, and while they are useful in comparing two populations that vary with one another, they are often not the best method of statistically interpreting bioimage data. Differences in shape or signal intensity between multiple samples can be analysed using analysis of variance (ANOVA) tests which show if there is a significance change in the population sampled. To determine which individual sample in the group analysed by ANOVA is different, post hoc testing of one or more metrics, e.g. the area of a nucleus or the elongation of an axon, can be carried out using Bonferroni or Dunnett's test. Bonferroni cross-correlates each group with the other groups to see which ones are significantly different from the others (many to many analysis). Dunnett's test allows a selected group to be compared with all of the others (one to many analysis). The Tukey HSD test is also useful for identifying arithmetic means of a group which are

Table 1.1 List softw	are for measuring and processing	bioimages.	
Image Analysis package	Open-source or manufacturer	Data types analysed or function	Reference link
Acapella	Perkin-Elmer	High content screens of 2D images, large volumes	http://www.perkinelmer.co.uk/ category/image-analysis- software
Amira+Avizo3D	FEI	Visualisation and quantification of 2D and 3D data, electron	https://www.fei.com/software/ avizo3d/
Arivis Vision 4D	Arivis	Modular software for working with multichannel 2D, 3D and 4D images of almost unlimited size independent of available RAM	https://www.arivis.com/en/ imaging-science/arivis- vision4d
BioimageXD	Open-source – Universities of Jyväskylä and Turku in Finland, MPI- CBG, Germany	3D quantification and segmentation of fluorescent images.	http://www.Bioimagexd.net/
Bisque	Open-source – UCSB, USA	Store, visualise, organise and analyse images in the cloud	http://Bioimage.ucsb.edu/ bisque
BigDataViewer	Open-source – MPI-CBG, Dresden Germany	BigDataViewer is a reslicing and quantifying browser for terabyte- sized 3D multiview image sequences	http://imagej.net/ BigDataViewer
Cellprofiler + Cellprofiler Analyst	Open-source – Broad Institute USA, [16]	Quantitatively measure phenotypes from many 2D images automatically and machine learning	http://cellprofiler.org/

Development of a fast and http://www.cellcognition.org/ cross-platform image analysis framework for fluorescence time- lapse microscopy in the field of bioimage informatics.	Accelerates development of http://www.definiens.com/ quantitative tissue-based image analysis. using machine learning	Flexible Java image processing http://imagej.net/Fiji/ program with a strong, Downloads established user base and thousands of plugins and macros for performing a wide variety of tasks	Software resources to visualise, http://icy.Bioimageanalysis.org/ annotate and quantify bioimage data, using a friendly user interface and coded in Python and Java.	3D and 4D real-time interactive http://www.bitplane.com/ data visualisation, quantification imaris/ and management	<i>ilastik</i> is a simple, user-friendly http://ilastik.org/ tool for interactive image classification, segmentation and analysis
Open-source – IMBA Vienna	Definiens	Open-source – Worldwide but led by Wayne Rasband (Ex NIH, USA) and Curtis Rueden, UW- Madison LOCI, USA) [16]	Open-source Institute Pasteur Paris	Bitplane, Oxford Instruments	Open-source – EMBL Heidelberg Germany and HHMI Janelia Farm USA
Cell cognition	Definiens	Fiji/ImageJ/ ImageJ 2	Icy	Imaris	Ilastik

lable 1.1 (Continue	(m)		
Image Analysis package	Open-source or manufacturer	Data types analysed or function	Reference link
ImageProPremier + Autoquant	MediaCybernetics	2D image analysis software to capture, process, count, classify, measure and share and also	http://www.mediacy.com/ imagepropremier
ImageTool	University of Texas USA	enables deconvolution Free C++ image processing and analysis program for pathology and fluorescence 2D images for Wish down and decome	http://compdent.uthscsa.edu/ dig/itdesc.html
Kalaimoscope	Open-source – MPI-CBG Dresden Germanv	windows platforms Object tracking in 2D and 3D	http://www.kalaimoscope. com/nroduct html
Leica LAS	Commercial, Leica Microsystems	Processing and quantification of 2D and 3D images acquired on Leica microscopes using Leica	http://www.leica- microsystems.com/home/
MetaMorph/ MetaExpress	Molecular Devices	acquisition software GPU accelerated image analysis of 2D, super-resolution and large volume datasets; some 3D image processing supported	https://www.moleculardevices. com/systems/metamorph- research-imaging/metamorph- microscopy-automation-and-
NIS Elements	Nikon Instruments	Processing and quantification of 2D, 3D and super-resolution images acquired on Nikon microscopes using Nikon acquisition software	image-analysis-software https://www. nikoninstruments.com/ en_GB/Products

BiG	Open-source – Bioimaging Group, EPFL Switzerland	Development of new algorithms and mathematical tools for advanced processing of biomedical images, e.g. SpotTracker, Deconvolution Lab	http://bigwww.epfl.ch/
TrackMate	Open-source – Institute Pasteur, France	TrackMate is your buddy for your everyday tracking of 2D and 3D objects	http://imagej.net/TrackMate
Vaa3D	Open-source – Allen Institute of Brain Science, USA	Explores big 3D/4D/5D images with giga-voxels, extracts complex surface objects from images and performs comprehensive analyses such as brain connectome mapping	http://www.alleninstitute.org/ what-we-do/brain-science/ research/open-science- research-tools/vaa3d/
Visiopharm	Visiopharm.	Quantitative digital pathology solution providing flexible, powerful quantification of tissue properties in a fast, objective and reproducible way	http://www.visiopharm.com/
Zen	Zeiss Microscopy	Processing and quantification of 2D and 3D images acquired on Zeiss microscopes using Zeiss acquisition software	http://www.zeiss.com/ microscopy/en_de/home. html

significantly different from one another because the means of several different samples can be tested. If the variance between multiple parameters is of interest, particularly in high content screens, principal component analysis can be particularly helpful. When comparing between populations it is important to bear in mind that many statistical tests assume that the data is normally distributed; this is true of t-tests, ANOVA and any post hoc tests associated with ANOVA. This type of test is called a parametric test. However, many bioimaging experiments, e.g. using CrispR knock-in or siRNA knock-out of genes will cause the distribution of phenotypes to become skewed or not normally distributed. This means that parametric statistical tests can no longer be used and non-parametric analysis based on the median or rank of data is more appropriate. Non-parametric tests which are commonly used in bioimage analysis are Mann-Whitney, which is the non-parametric analogue of the t-test or Kruskal–Wallis which is a non-parametric analogue of ANOVA. An in-depth description of statistics is beyond the scope of this book, but some the following sources of information are useful.

- Nature Statistics for Biologists collection. http://www.nature.com/ collections/qghhqm
- http://www.wormbook.org/chapters/www\_statisticalanalysis/ statisticalanalysis.html
- Statistical and Data Handling Skills in Biology, Roland Ennos, Pearson Education Ltd ISBN: 9780273729495
- Statistics for Terrified Biologists, Helmut van Emden, Wiley ISBN: 978-1-4051-4956-3

Statistical analysis tools for data mining and informatics analysis that can be applied to large multivariate datasets, such as high content screening include: 'Analytics' applications such as the open-source Cellprofiler Analyst, Bisque and KNIME (https://www.knime.org/) or commercially available SpotFIRE (http://spotfire.tibco.com/), AcuityXpress from molecular devices and others that have toolboxes to enable multiparametric analysis, data visualisation and pipelines into comparative analysis with genomic, transcriptomic and proteomic analysis datasets. It is more routine, for smaller datasets to prepare the data for presentation at this point in the experimental process. Strategies for this are discussed in Chapter 10.

Ann Wheeler would like to thank Dr Ricardo Henriques and the IGMM advanced imaging facility users for constructive comment on this chapter.

## **1.5 REFERENCES**

- Spector, D. and R. Goldman, 2005. Basic Methods in Microscopy Protocols and Concepts from Cells: A Laboratory Manual. s.l.: Cold Spring Harbor Press ISBN 978-087969751-8.
- Husz, Z.L., N. Burton, B. Hill, N. Milyaev and R.A. Baldock, 2012. Web tools for large-scale 3D biological images and atlases., BMC Bioinformatics., pp. 13:122. doi: 10.1186/1471-2105-13-122.
- Sternberg, S.R., 1983. *Biomedical Image Processing*, IEEE Computer, pp. Volume: 16, Issue: 1, 22–34, 10.1109/MC.1983.1654163.
- 4. Swedlow, J.R., 2003. *Quantitative fluorescence microscopy and image deconvolution*. Methods Cell Biol., pp. 72:349–67. PMID:14719340.
- Richardson, W.H., 1972. Bayesian-Based Iterative Method of Image Restoration. Journal of the Optical Society of America, pp. 62 (1): 55–59. doi:10.1364/JOSA.62.000055.
- Lucy, L.B., 1974. An iterative technique for the rectification of observed distributions, Astronomical Journal., pp. 79 (6): 745–754. Bibcode:1974AJ...79...745 L. doi:10.1086/111605.
- 7. Lane, R.G., 1996. *Methods for maximum-likelihood deconvolution*. Journal of the Optical Society of America, pp. vol. 13, Issue 10, pp. 1992–1998 doi. org/10.1364/JOSA.13.001992.
- 8. Zitova, B. and J. Flusser, 2003. *Image Registration Methods, a Survey*. Image and Vision Computing, pp. 21, 21977–1000.
- Wang C.-W., Ka S.-M. and A. Chen, 2014, Robust image registration of biological microscopic images. Scientific Reports 4, Article number: 6050, p. doi:10.1038/srep06050.
- Alvaro J.P. and M.B. Arrate, 2015, Free Form Deformation Based Image Registration Improves Accuracy of Traction Force Microscopy. Plos One, p. 10.1371/journal.pone.0144184.
- 11. Cardona A., S. Saalfeld, J. Schindelin, et al., 2012, *TrakEM2 Software for Neural Circuit Reconstruction*.PLoS ONE, p. 10.1371/journal.pone.0038011.
- 12. Ess, A., T. Tutelaars and L. Gool, 2008, *Speeded Up Robust Features (SURF)*. Comput. Vis. Imag. Understand., pp. 110, 346–359.
- Wang, C. and H. Chen, 2013, SIFT Improved Image Alignment Method in application to X-ray Images and Biological ImagesBioinfo., pp. 29, 1879–1887.
- 14. Peng, H., Z. Ruan, F. Long, J.H. Simpson and E.W. Myers, 2010. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nat. Biotechnology, pp. 28, 348–53 doi:10.1038/nbt.1612.
- 15. Peng, H., P. Chung, F. Long et al., 2011, *BrainAligner: 3D registration atlases* of *Drosophila brains*. Nat. Methods, pp. 8, 493–498 doi:10.1038/nmeth.1602.
- Schneider, C.A., W.S. Rasband and K.W. Eliceiri, 2012, NIH Image to ImageJ: 25 years of image analysis. Nature Methods, pp. 9, 671–675 doi:10.1038/ nmeth.2089.