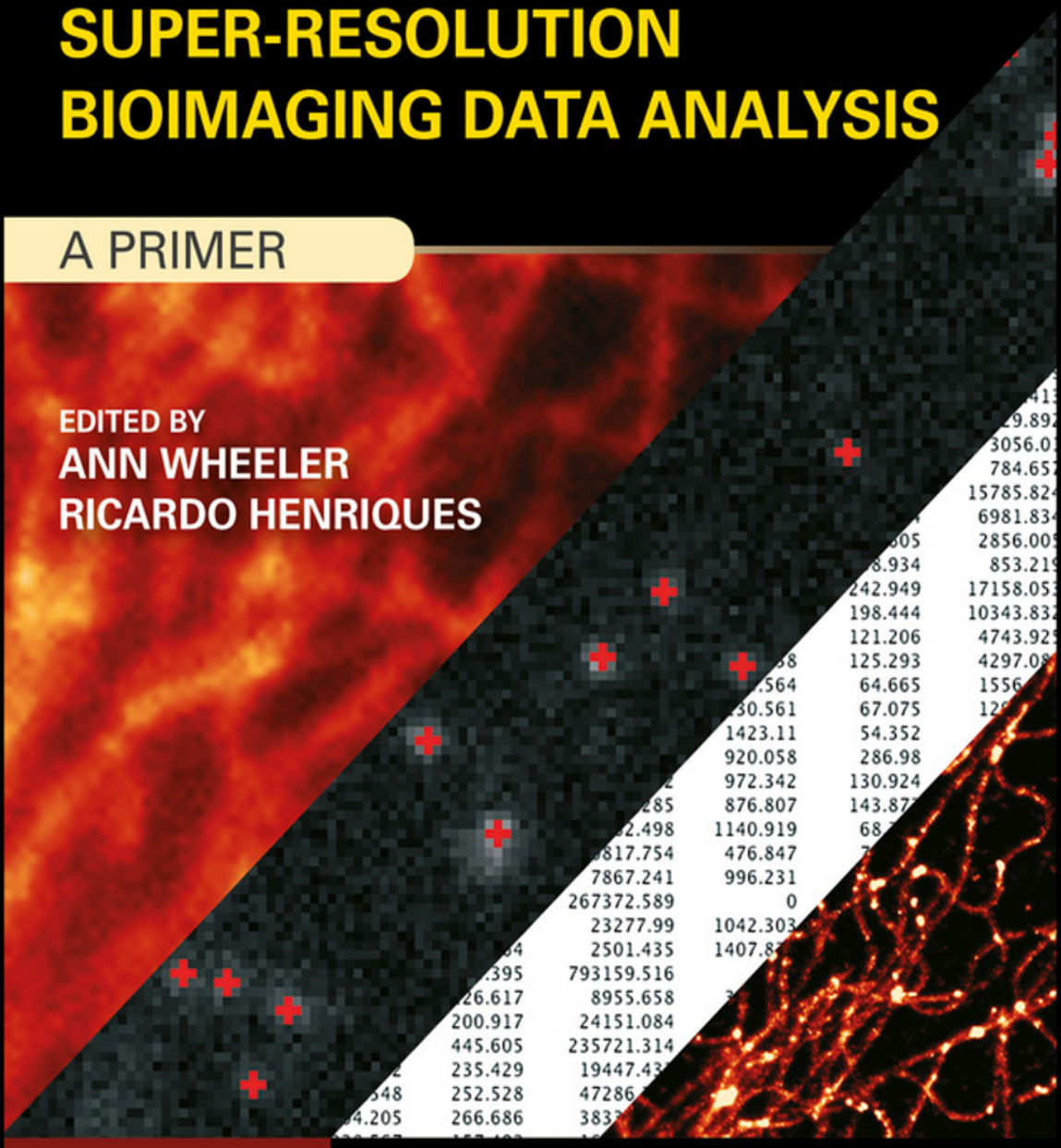
STANDARD AND SUPER-RESOLUTION BIOIMAGING DATA ANALYSIS

A PRIMER

EDITED BY
ANN WHEELER
RICARDO HENRIQUES



RMS

WILEY

Standard and Super-Resolution Bioimaging Data Analysis

Current and future titles in the Royal Microscopical Society—John Wiley Series

Published

Principles and Practice of Variable Pressure/Environmental Scanning Electron Microscopy (VP-ESEM)
Debbie Stokes

Aberration-Corrected Analytical Electron Microscopy
Edited by Rik Brydson

Diagnostic Electron Microscopy—A Practical Guide to Interpretation and Technique
Edited by John W. Stirling, Alan Curry & Brian Eyden

Low Voltage Electron Microscopy—Principles and Applications
Edited by David C. Bell & Natasha Erdman

Standard and Super-Resolution Bioimaging Data Analysis: A Primer
Edited by Ann Wheeler and Ricardo Henriques

Forthcoming

Understanding Practical Light Microscopy
Jeremy Sanderson

Atlas of Images and Spectra for Electron Microscopists
Edited by Ursel Bangert

Focused Ion Beam Instrumentation: Techniques and Applications
Dudley Finch & Alexander Buxbaum

Electron Beam-Specimen Interactions and Applications in Microscopy
Budhika Mendis

Standard and Super-Resolution Bioimaging Data Analysis: A Primer

Edited by

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

Ricardo Henriques
MRC Laboratory for Molecular Cell Biology
University College London, UK

Published in association with the Royal Microscopical Society

Series Editor: Susan Brooks

RMS

WILEY

This edition first published 2018

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Wheeler, Ann, 1977– editor. | Henriques, Ricardo, 1980– editor.
Title: Standard and Super-Resolution Bioimaging Data Analysis: A Primer / edited by Dr. Ann Wheeler, Dr. Ricardo Henriques.
Description: First edition. | Hoboken, NJ : John Wiley & Sons, 2018. | Includes index. |
Identifiers: LCCN 2017018827 (print) | LCCN 2017040983 (ebook) | ISBN 9781119096924 (pdf) | ISBN 9781119096931 (epub) | ISBN 9781119096900 (cloth)
Subjects: LCSH: Imaging systems in biology. | Image analysis–Data processing. | Diagnostic imaging–Data processing.
Classification: LCC R857.O6 (ebook) | LCC R857.O6 S73 2017 (print) | DDC 616.07/54–dc23
LC record available at https://lccn.loc.gov/2017018827

Cover design by Wiley
Cover image: Courtesy of Ricardo Henriques and Siân Culley at University College London

Set in 10.5/13pt Sabon by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

List of Contributors

George Ashdown, Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, UK

Graeme Ball, Dundee Imaging Facility, School of Life Sciences, University of Dundee, UK

Sébastien Besson, Centre for Gene Regulation & Expression and Division of Computational Biology, University of Dundee, UK

Mario De Piano, Division of Cancer Studies, King's College London, UK

Ahmed Fetit, Advanced Imaging Resource, MRC-IGMM, University of Edinburgh, UK
and
School of Science and Engineering, University of Dundee, UK

Juliette Griffié, Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, UK

Aliaksandr Halavatyi, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Ricardo Henriques, MRC Laboratory for Molecular Cell Biology, University College London, UK

Gareth E. Jones, Randall Division of Cell & Molecular Biophysics, King's College London, UK

Debora Keller, Facility for Imaging by Light Microscopy, Imperial College London, UK

Kota Miura, Nikon Imaging Center, Bioquant, University of Heidelberg, Germany; National Institute of Basic Biology, Okazaki, Japan; Network of European Bioimage Analysts (NEUBIAS)

Nicolas Olivier, Department of Physics and Astronomy, University of Sheffield, UK

Peter O'Toole, Technology Facility, Department of Biology, University of York, UK

Dylan Owen, Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, UK

Thomas Pengo, University of Minnesota Informatics Institute, University of Minnesota Twin Cities, USA

Michael Shannon, Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, UK

Stefan Terjung, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Jean-Yves Tinevez, Institut Pasteur, Photonic BioImaging (UTechS PBI, Imagopole), Paris, France

Sébastien Tosi, Advanced Digital Microscopy Core Facility (ADMCF), Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology, Barcelona, Spain; Network of European Bioimage Analysts (NEUBIAS)

Claire M. Wells, School of Cancer and Pharmaceutical Sciences, King's College London, UK

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

Foreword

Imaging is now one of the most commonly used techniques in biological research. It is not simply a means of taking a pretty picture; rather advanced microscopy and imaging are now vital biophysical tools underpinning our studies of the most complex biological systems. We have the capability to study cells in real time, with 3D volumes, analysing biophysical interactions, and now moving towards the ability to see and study individual proteins within individual cells within their native environment. Imaging is one of the most useful tools for understanding the fundamental biology of the cell.

This has been made possible through an incredibly rapid period of microscopy development, which has gone hand in hand with the emergence of new fluorescent tags – such as fluorescent proteins – computing power and the latest developments in engineering. Not only has the technology become increasingly versatile and opened up many new possibilities, but leading manufacturers have also made their microscopes increasingly accessible to non-specialists, resulting in an explosion of data.

All of these developments have left us with a wealth of data, but the images themselves will remain just pretty pictures unless they are analysed appropriately. We are now starting to see an equivalent rapid increase in the development of image analysis, but we are still far from realising its full potential. The basics are vital, and anyone using today's microscopes should also be looking at the best approaches for analysing their data.

This book is extremely timely and looks at some of the key aspects of data analysis; it will serve as an excellent point of reference. Chapter 1 examines the basics of image data and processing which is common to most users. Chapters 2 and 3 builds on this and looks at how to quantify both routine 2D through to more complex 3D image data sets. However, one of the greatest challenges remains the ability to segment our images. To our own eyes, this can be quite obvious, but it still remains a real computational challenge. Segmenting images and image data

e.g. in Chapter 3 will be a continuing area of development that will also help us to correlate data with greater precision in the future.

Beyond the images, the microscope is a powerful biophysical tool. We can see and analyse the co-localisation of particles such as proteins, but great care is needed in their analyses as outlined by Dylan Owen e.g. in Chapter 6. We can now go beyond this and start to see interactions that occur over a 5 nm range. This enables the studies of particle–particle interactions such as protein–protein heterodimerisation by using FRET, and although the imaging of these phenomena is relative simple the complexity of the quantification is discussed in Chapter 4.

Not only can the microscope study these natural interactions, but we can also use the light, often with lasers, to manipulate cells and trigger critical events that would otherwise occur in a random fashion making their studies very difficult. The interpretation and controls for FRAP and other photo perturbation methods then needs to be carefully considered (Chapter 5).

Many of the above studies can be undertaken using both fixed and live cell imaging. Whole live-cell imaging and tracking brings its own analytical challenges (Chapter 7) as cells move through three dimensions, pass over one another often changing shape and nature. This needs many of the above elements to come together to help work in such complex sample types.

At the other extreme, from whole cells, the biggest advancement in light microscopy has come from the ability to image below the diffraction limit. Super-resolution microscopy (SRM) is possible through many different strategies. The analysis and interpretation is an area that is often under-appreciated and which can result in misinterpretations. For anyone wanting to undertake SRM, it is essential to understand the limitations, controls and best approaches to their analysis (Chapter 8).

Many of the new microscopical techniques now produce very large data sets. This is especially true for 3D live-cell imaging and SRM. This has developed its own problem, with data analyses now often taking considerably longer than the imaging time itself. The time needed for data analysis has become the most costly element in complex image study. The staff time itself often outweighs the cost of the instrument and consumables and this is why we need to look for automation when handling and analysing the data (Chapter 9), and naturally, once all of this data has been analysed, it is vital to not only present the data in the correct manner, but also to ensure that it is correctly documented and stored (Chapter 10). Only then, can any one image be properly exploited and deliver the required impact.

Peter O'Toole
York
July 2017

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.

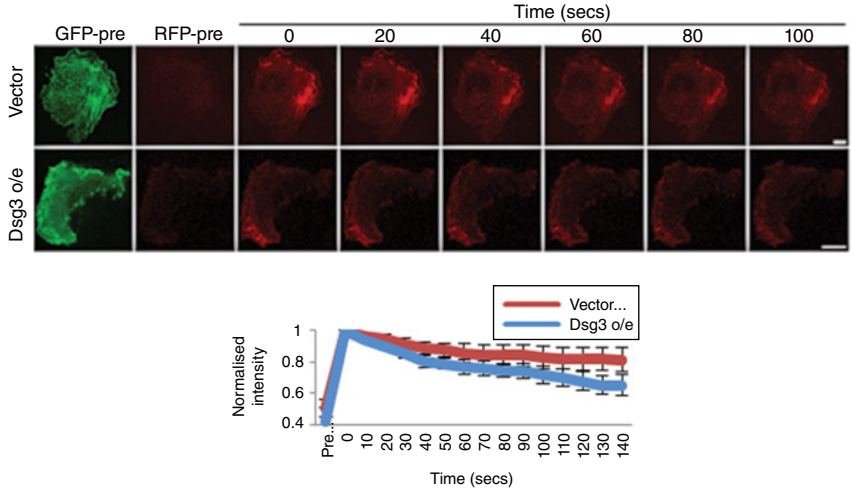


Figure 1.1 Bioimage quantification to determine the dynamics of actin using photoconversion. Tsang, Wheeler and Wan Experimental Cell Research, vol. 318, no. 18, 01.11.2012, p. 2269–83.

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).

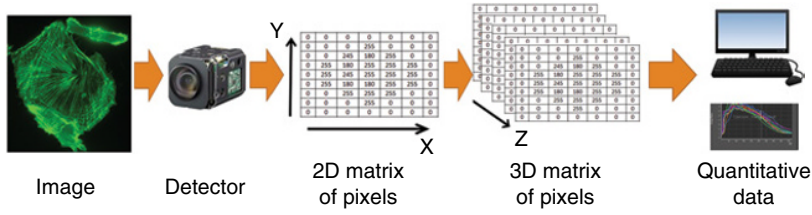


Figure 1.2 Workflow for bioimage data capture in 2D and 3D.

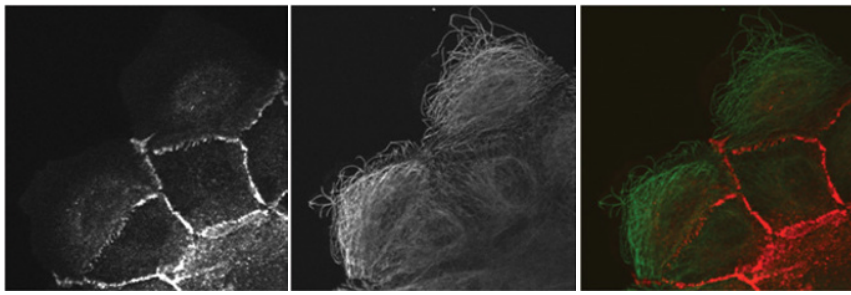


Figure 1.3 Combining channels in fluorescent bioimage analysis. Channel 1 has antibodies raised against E-cadherin labelled with AlexaFluor 568 secondary antibodies. Channel 2 is labelled with primary antibodies raised against Alpha tubulin and secondary antibodies labelled with AlexaFluor 488.

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 analysis needs to be used to analyse the differences in the biological system.

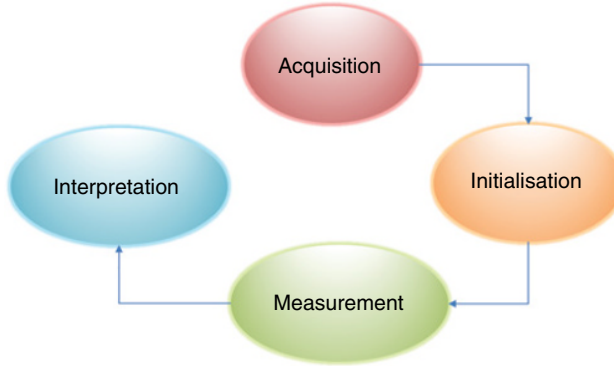


Figure 1.4 The Bioimage analysis workflow.

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

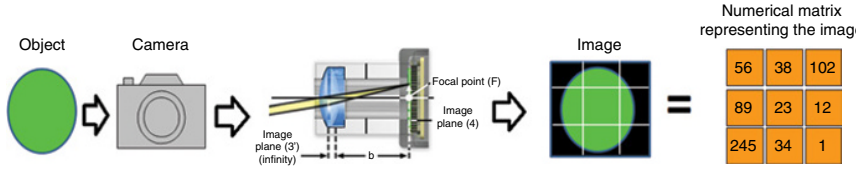


Figure 1.5 How images are digitised.

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)

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 2^8, 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).

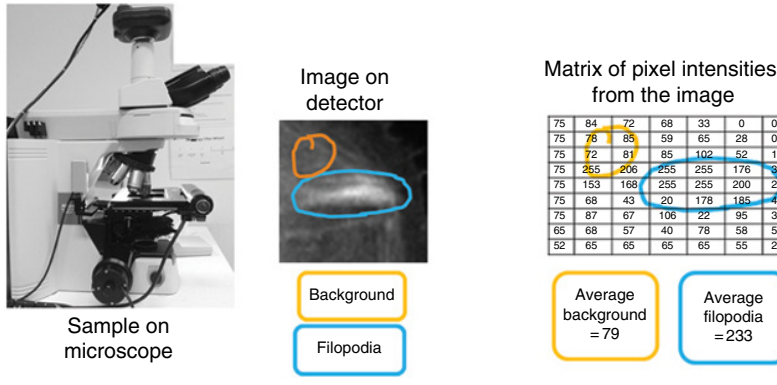


Figure 1.6 Basic quantification of cellular features using 8-bit fluorescent image of F-actin.

It is possible to increase the sensitivity of the pixel further, currently to 12 (4096 or 2^{12}), 14 (16384 or 2^{14}) and 16 (65536 or 2^{16}) 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^{16} 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

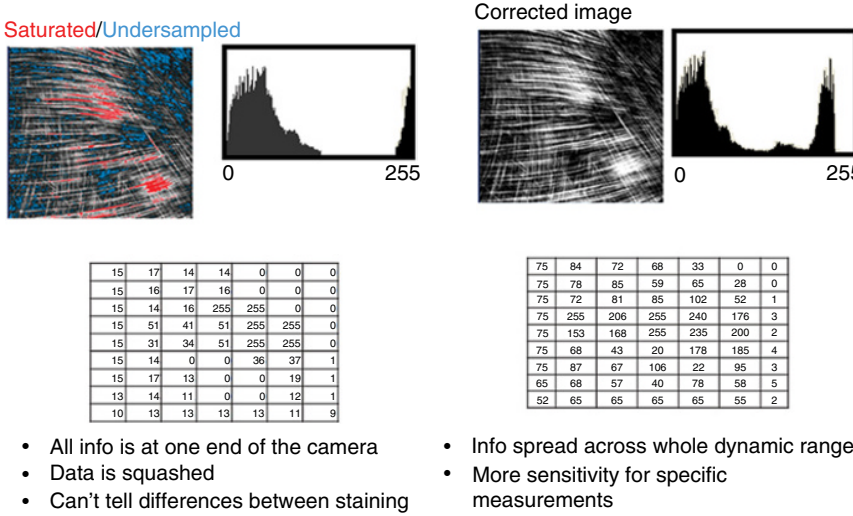


Figure 1.7 The effect of saturation and under-sampling on bioimage analysis.

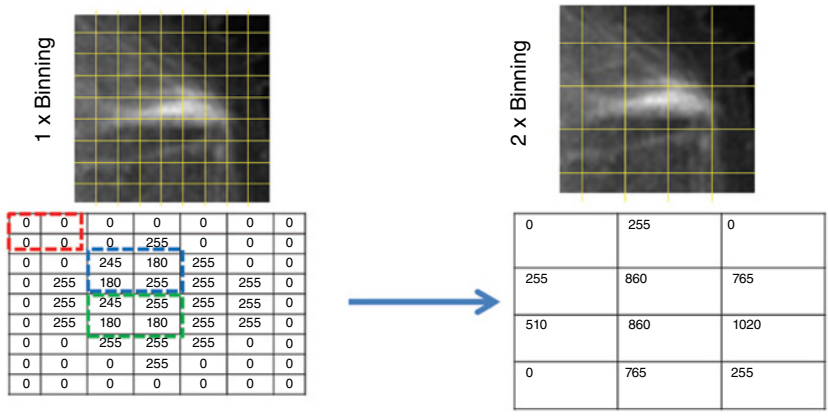
instance if imaging a very fast process or when a very weak signal is being collected from a probe which can be photo-damaged. Provided that sufficient signal can be collected for quantitative analysis this need not be a problem. However, best practice is to have the signal fill the whole dynamic range 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 2×2 array of pixels or a 4×4 array is combined into one super-pixel. The advantage of this is that there is a 4 (2×2 bin) or 16 (4×4) 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.



Advantages	Disadvantages
More sensitivity	Less resolution
Data collected faster	Data is less detailed
Calcium imaging / FRET	Migration and localisation assays

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).

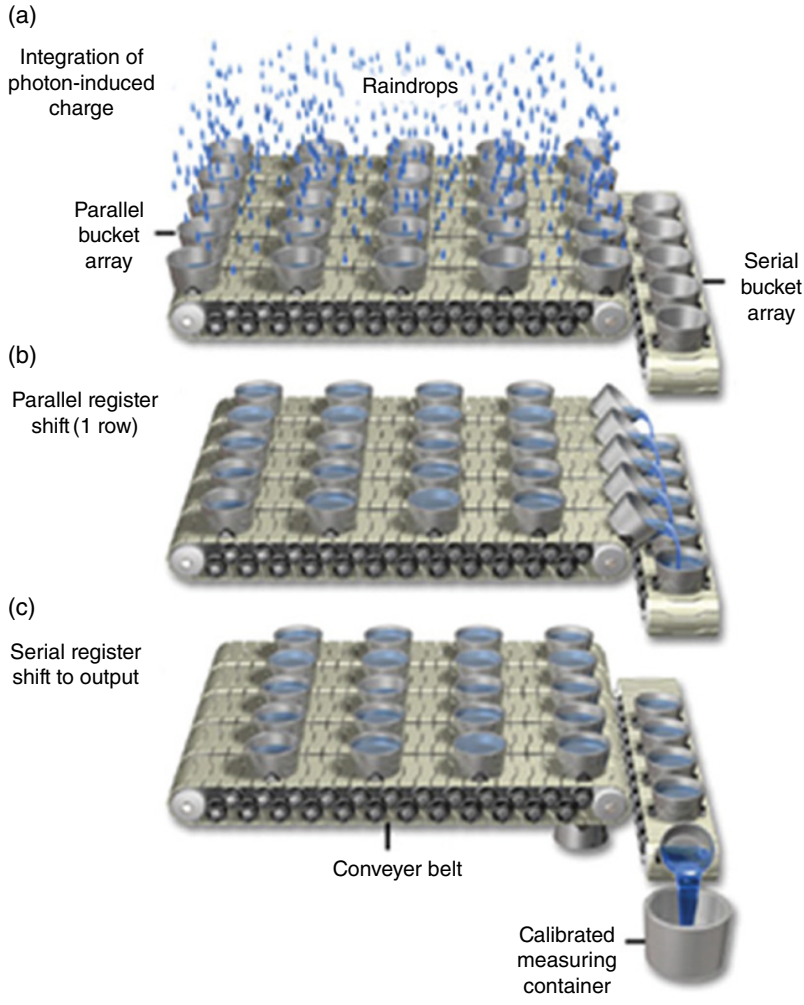


Figure 1.9 Bucket brigade CCD analogy (Courtesy of Molecular Expressions, Florida state Univeristy, USA, https://micro.magnet.fsu.edu/primer/index.html).

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

Median = 79

Figure 1.10 A 3 × 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.