Toward the Automated Detection and Characterization of Osteoclasts in Microscopic Images

2

Andreas Heindl, Martin Schepelmann, Robert Nica, Rupert Ecker, Peter Pietschmann, Alexander K. Seewald, Theresia Thalhammer, and Isabella Ellinger

2.1 Introduction

Microscopes have been used for a long time to observe biological samples. However, measurements of tissue- and cell-related parameters were conducted by human observers and were consequently ad hoc, not reproducible and restricted to small sample numbers. Since computers have become vastly more powerful, life sciences now routinely take advantage of new opportunities to couple microscopy and in silico methods. Automated image segmentation and analysis of large numbers of digital images allow algorithmic recognition of cell and tissue structures and subsequent numeric measurements of cellular parameters. Nevertheless, these new

A. Heindl, PhD (✉)
Department of Pathophysiology and Allergy Research, Center for Pathophysiology, Infectiology and Immunology, Medical University of Vienna, Vienna, Austria

Seewald Solutions, Weisskirchen an der Traun, Austria

Division of Molecular Pathology and Centre for Evolution and Cancer, The Institute of Cancer Research, 15 Cotswold Road, London SM2 5NG, UK
e-mail: andreas.heindl@icr.ac.uk

M. Schepelmann, PhD
Department of Pathophysiology and Allergy Research, Center for Pathophysiology, Infectiology and Immunology, Medical University of Vienna, Vienna, Austria

School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, UK

R. Nica • R. Ecker, PhD
TissueGnostics GmbH, Vienna, Austria

P. Pietschmann, MD • T. Thalhammer, PhD • I. Ellinger, PhD
Department of Pathophysiology and Allergy Research, Center for Pathophysiology, Infectiology and Immunology, Medical University of Vienna, Vienna, Austria

A.K. Seewald, PhD
Seewald Solutions, Weißkirchen an der Traun, Austria

P. Pietschmann (ed.), *Principles of Osteoimmunology*,
DOI 10.1007/978-3-319-34238-2_2

methods also come with technical challenges concerning computational resources like processing capacity, memory and disk space, biological sensor limitations, as well as algorithm development.

Today, state-of-the-art hardware and cloud computing enable high throughput analysis of vast amount of images in reasonable time. In the context of bone research, an application for automated analysis is the in silico quantification of osteoclasts. Cell culture models using either murine or human osteoclasts offer the possibility to study parameters such as osteoclast formation from their mesenchymal precursors, differentiation, maturation, and apoptosis (Marino et al. 2014). Advanced molecular imaging of osteoclasts allows to study pathological processes and to elucidate the effect of osteoclast-targeted therapies for diseases in which excess bone resorption is a crucial pathological process. This includes osteoporosis, rheumatoid arthritis, as well as bone tumors such as giant cells tumors, osteosarcomas, and bone metastases.

Currently, quantification of multinucleated osteoclasts in culture is performed manually (Marino et al. 2014). When studying the influence of endogenous metabolites, hormones or therapeutic agents on osteoclast biology, a lot of potentially valuable and relevant information, such as number of nuclei, osteoclast size, number and properties of precursor cells, as well as abundance of target proteins in each cell class, cannot be assessed reliably by manual evaluation. These parameters, however, could be important to elucidate the effects of natural and synthetic agents on osteoclast biology.

To exemplify a typical question in basic osteoclast research, we were looking at whether the pineal hormone melatonin affects osteoclast formation in vitro. Melatonin has long been sought to exert beneficial effects on bone structure and has been proposed to prevent osteoporosis in premenopausal and menopausal women (reviewed in Maria and Witt-Enderby (2014)). Melatonin was found to favor the differentiation of human adult's mesenchymal stem cell into osteoblasts via binding and signaling through a G-protein-coupled melatonin receptor (Radio et al. 2006; Koyama et al. 2002). Thus, beneficial effects of melatonin on bone structure were mainly attributed to its effects on osteoblasts (Maria and Witt-Enderby 2014; Radio et al. 2006). Although melatonin may prevent bone degradation by inhibiting osteoclast formation directly (Ostrowska et al. 2001), the inhibitory effects on osteoclast are also caused by the action of melatonin on osteoblasts. This was suggested from studies in mice, where melatonin increased the osteoprotegerin expression in osteoblasts, which lead to a reduction in the RANKL-mediated osteoclast formation (Maria and Witt-Enderby 2014; Koyama et al. 2002). In a pilot experiment, we re-investigated the possibility of direct effects of melatonin on the formation of mouse osteoclasts from bone-derived precursor cells specifically by focusing on osteoclast number, cell area, multinuclearity, and downregulation of the macrophage (osteoclast precursor cell) marker protein F4/80 as osteoclast marker. As outlined above, limitations of manual evaluation of osteoclast formation would normally prevent the detection of small-scale correlations between changes of cell-associated parameters and the effects of the hormone melatonin. However, when employing a computerized quantification method, all of the abovementioned limitations can be overcome.

In this chapter, we briefly describe standard procedures for osteoclast culture, markers for osteoclast detection, as well as the applied immunofluorescence labeling protocol. We provide a brief introduction to digital images and slide-based microscopy and discuss software packages for image processing as well as valuable tools for handling the workflow.

We then describe the development and application of an image processing algorithm for the detection and quantification of osteoclasts. Principles of ground-truth data and their evaluation as well as the algorithms used for osteoclast detection are explained. Measurements are introduced, and examples are given on how to apply them in real-life scenarios. The described steps are not programming language specific and can be implemented in the framework of the reader's choice. However, we do not cover implementation details, as these are beyond the scope of this work and should be read in existing books on microscopy and digital image processing, for example (Gonzalez and Woods 2008; Wu et al. 2008; Burger and Burge 2008; Solomon and Breckon 2011). The algorithm, which was developed, has been integrated in the software application StrataQuest (TissueGnostics GmbH, Austria), a context-based qualitative image analysis software package. As a "real-life" basic research example, we then show the results of our pilot study using StrataQuest to quantify the direct effect of melatonin on osteoclast formation in culture which yielded novel interesting results that would not have been available without an automated in silico analysis system.

Many of the image processing problems mentioned in these sections are so-called *ill-defined* problems – e.g., segmentation (Martin et al. 2001; Bakushinskiy and Goncharsky 2012) – meaning that there is no unambiguous gold standard to compare these algorithms to. Furthermore, the question of how to compare a developed algorithm to the performance of human experts still remains an important open research problem. Thus, to illustrate these difficulties, we cover not only typical pitfalls including technical problems but also human intuition and limitations in perception and vision that have an influence on the development and evaluation of image analysis algorithms.

In summary, the purpose of this chapter is to introduce biologists and medical scientists to image processing by using a commercially available automated image analysis program to quantify various parameters in osteoclast cultures and exemplify this through evaluating the effect of melatonin on osteoclast formation. We hope that we can thereby raise the audience's awareness and interest in the possibilities and limitations of this new, powerful technology.

2.2 Techniques

2.2.1 Culture Conditions for Isolated Murine Osteoclasts

The culture conditions for osteoclasts suitable for an automatic detection method were developed from standard protocols (Marino et al. 2014; Akatsu et al. 1992), but different culture parameters were evaluated.

Mice (*Mus musculus*) were culled by neck dislocation following asphyxiation. Tibiae and femora were prepared, the caps of the bones were cut off, and the bones were rinsed with 10 ml pre-warmed (37 °C) minimum essential medium containing antibiotics and antifungals. The cells were diluted to 2×10^6 cells per ml, and osteoclast formation-stimulating additives (25 ng/ml Receptor activator of nuclear factor kappa-B ligand (RANKL), 15 ng/ml Macrophage colony-stimulating factor (M-CSF)) were added. 1 ml of cell suspension was added per well (each containing a sterile glass coverslip) of a 24-well culture plate. Culture was maintained at 95 % relative humidity/5 % CO_2 in an incubator for 7 days. Change of medium was performed every second day.

To investigate the effects of melatonin on osteoclast development from bone marrow precursor cells, we applied different doses of melatonin (1, 0.1, and 0.01 μM melatonin from a stock solution of 10 mg/1 ml Dimethyl sulfoxide (DMSO)) to the osteoclast cultures for the entire cultivation time. The respective controls were treated with the solvent only.

2.2.2 Staining Protocol

At the beginning of the development of an algorithm, characteristic features of the target structure or cell need to be defined by biological experts, preferably in a written document for later reference. In the case of osteoclasts, the biological experts defined two important criteria to identify mature osteoclasts: *criterion 1,* the amount of nuclei per cell (≥3) (Andersson and Marks 1989) and *criterion 2*, a low to undetectable expression of the macrophage-antigen F4/80 (van de Wijngaert et al. 1987). The expression of the latter marker is reduced or lost due to the differentiation from precursor cells to osteoclasts. Therefore, the staining protocol included (1) labeling of the nuclei with blue DAPI, (2) staining of the cells with an antibody directed against F4/80 macrophage marker (eBioscience.com), probed with red Alexa Fluor 568 (Invitrogen molecular probes), and (3) making all cells (osteoclasts and precursors alike) "visible" using one antibody against the membrane-bound calcitonin receptor (Acris) and one antibody against the cytoskeleton component α-tubulin (Sigma Aldrich), both probed with Alexa Fluor 647 (Invitrogen molecular probes; far red, appearing white in the image). In the acquired images (e.g., Fig. 2.1), mature osteoclasts appear white (due to the staining for α-tubulin and reduced expression of F4/80 staining), while precursor cells appear red or pink from the α-tubulin – calcitonin receptor – F4/80 overlay. It should be kept in mind however that differentiation and fusion of cells to generate mature osteoclasts is a continuous process. Therefore, mononucleated cells with rather low F4/80 staining can be present in the culture (visible in Fig. 2.1) as well as multinucleated cells with higher expression levels of F4/80. It is up to the biological experts to define a threshold level of mean F4/80 intensity/cell that discriminates between "true" osteoclasts (≥3 nuclei per cell, mean fluorescence intensity of F4/80 < set threshold) and "true" precursor cells (1–2 nuclei/cell, mean fluorescence intensity of F4/80 > set threshold).

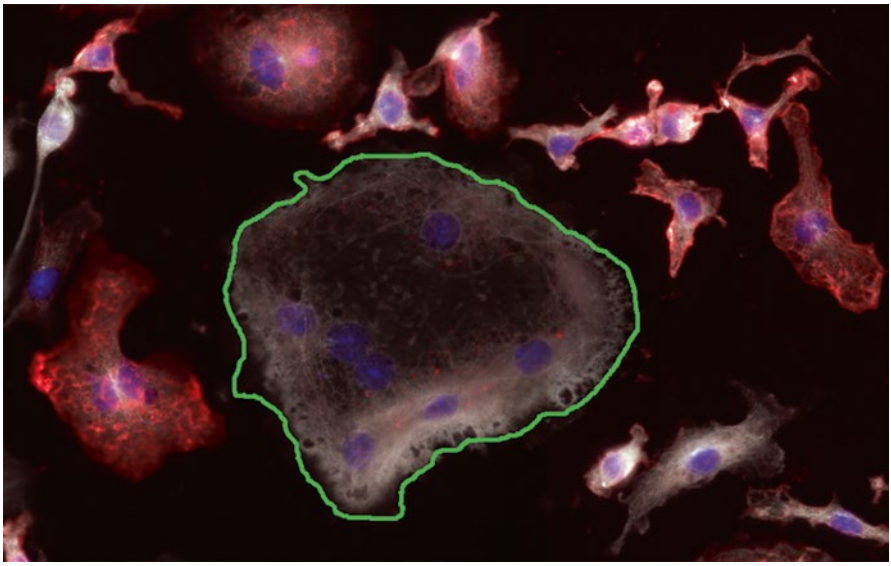

Fig. 2.1 This image shows one immunofluorescence-labeled osteoclast cell in culture surrounded by precursor cells. The osteoclast is marked up in *green*. The perimeter of the target cell is created on a separate layer so that the original content of the image is not modified. This additional layer can later be extracted and processed by computer-based algorithms

Cells were fixed using a 4 % formaldehyde solution, and the remaining aldehyde groups were quenched with 50 mm NH_4Cl. The cells were incubated with blocking/permeabilization buffer (0.5 % Triton X-100 + 1 % bovine serum albumin in phosphate-buffered saline (PBS)) for 60 min. After this period, the primary antibodies were applied directly in the culture plate at a dilution of 1:1,000 in blocking buffer (parallel approach) for 60 min. After washing with PBS, the secondary antibodies were applied in the dark at a dilution of 1:1,000 in blocking buffer (parallel approach) for 30 min. To stain the nuclei, the cells were incubated for 15 min with DAPI (1 µg/ml) in *aqua bidestilata*. The coverslips were finally mounted with Fluoromount-G (SouthernBiotech) on conventional microscope slides.

Images of the stained cells were acquired using an automated Axio Imager epifluorescence microscope (Zeiss) equipped with TissueFAXS™ hardware and software (TissueGnostics GmbH, Austria) using a 40× Neofluar 1.4 (oil) objective. For special considerations concerning the acquisition process, see the following sections.

2.2.3 Digital Images

Before algorithm development can commence, images have to be acquired and stored. The digitalization of specimens requires a suitable representation of the captured light by an electronic sensor (camera) that can be further processed by the

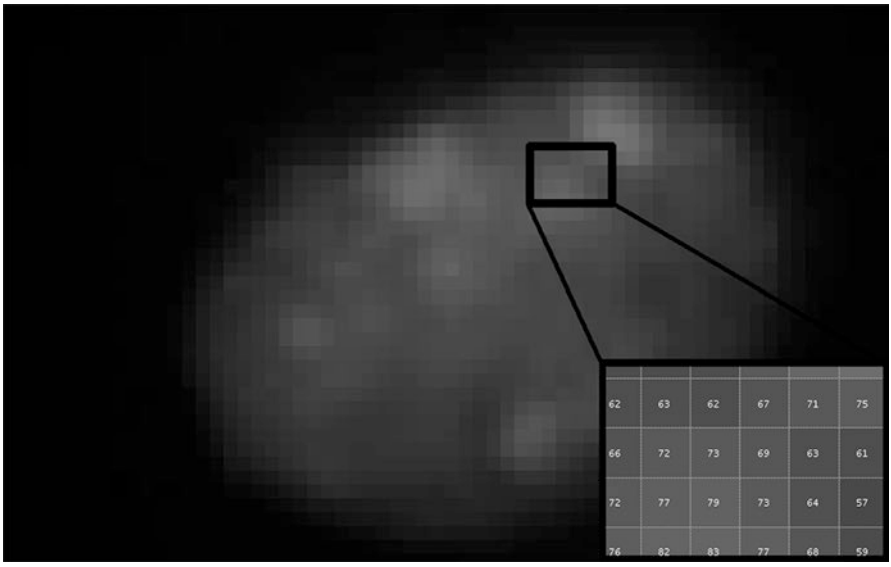

Fig. 2.2 This figure illustrates the digital representation of a captured image derived from fluorescence microscopy. The small insert on the right side shows the gray level value of the selected pixel (8-bit image: a value between 0 and 255)

computer. Exposure time controls the amount of light that is able to hit the sensor elements. Finally, quantification transforms these values to a limited set of intensities that can be further processed by the computer. Typical ranges of 256 ($=2^8=8$ Bit), 4096 ($=2^{12}=12$ Bit), or 65,536 ($=2^{16}=16$ Bit) are used to represent these intensity values. An example of an 8-bit image with a corresponding gray-level matrix is shown in Fig. 2.2.

2.2.4 Automated Slide-Based Microscopy

Sampling of images from large-scale experiments necessitates automation of the acquisition process. A suitable microscope with a fully motorized stage has to be used to automatically acquire images. During acquisition, the stage moves the slides in such a way that the camera captures each region of interest as a set of overlapping field of views (FOVs). For this image acquisition, we used a TissueFAXS™ system (TissueGnostics GmbH, Austria), which offers a convenient workflow to acquire up to eight slides automatically. This was a requirement for further statistical analysis to analyze global effects of various compounds on the growth and formation of osteoclasts in culture. If these cultures contain huge cells, in this case osteoclasts, then stitching (tiling together of the single FOVs to one big image) of acquired regions is essential for further steps, as these cells may have been only partially captured at the border of a FOV. Automatic-stitching algorithms have two main components: finding the best alignment for two neighboring images and finding the

best alignment taking into account all neighboring images. TissueFAXS already includes stitching during and after acquisition.

2.2.5 Software for Image Processing

Commercial state-of-the-art solutions like StrataQuest (TissueGnostics GmbH, Austria), in which the osteoclast detection algorithm we have developed was finally incorporated, can automatically detect tissue structures on a digital slide by integrating the detection of objects into detailed context-based quantitative analysis (Stadler et al. 2015; Schmid et al. 2015). StrataQuest allows the analysis of interactions between different types of data like nuclei counts, cell area, protein staining intensities, and the different cell types present via an easy-to-use graphical user interface (GUI). Coupled with TissueFAXS, acquisition and analysis are possible in a homogeneous workflow and CE-marked analysis environment for use in research and in vitro diagnostics. An open source alternative, which requires only basic knowledge of image processing, is Cell Profiler.[1] This application suite offers a framework to build versatile *pipes* (algorithms) for various biological applications. Many predefined pipes for various applications can be downloaded from the online forum and are free of charge. For immunohistochemically stained images, GemIdent[2] (Holmes et al. 2009) is also a versatile tool. A good overview of available software packages is provided in Eliceiri et al. (2012).

However, when designing a new algorithm, a developing framework has to be chosen. Common tools like GNU Image Manipulation Program (GIMP) or Adobe Photoshop are not suitable for algorithm development due to lack of a powerful and fast scripting language. High-performance algorithms made image software tools like OpenCV,[3] ImageJ (Rasband 1997), and Matlab[4] however are very popular: They offer special image-processing toolboxes that are optimized for high throughput while still being relatively easy-to-use for rapid prototyping of image analysis algorithms. For rapid prototyping of image analysis problems, Matlab and ImageJ are good choices. The disadvantage of Matlab and ImageJ is that they are computationally slow compared to OpenCV – so if speed is an issue, algorithms have to be developed or ported to this framework.

In any case, independent of the used framework, generation of ground-truth markups should be started as early as possible so that there is a continuous flow of data. Ground-truth data represent a set of images provided by biological experts that include counts, location, and key features of the objects of interest (Krig 2014). This dataset is then used to evaluate the newly developed algorithms, train machine-learning systems, and optimize parameters throughout the whole development process.

[1] http://cellprofiler.org

[2] http://gemident.com

[3] http://opencv.org/

[4] www.mathworks.com/

2.3 Evaluation of Expert Markups and Developed Image-Processing Algorithms

After and during development, the performance of newly implemented algorithms needs to be measured in an unbiased way. For this purpose, visual inspection is the most commonly used method. Unfortunately, this step is error prone and highly dependent on the observer who in many cases will still be a computer scientist rather than a qualified biological researcher. A better method for this evaluation is to let biological domain experts do markups to obtain the so-called ground-truth data from original images (Krig 2014). This can be done using any graphics-editing program like GIMP or Adobe Photoshop. The idea of this markup is to point out those objects of interest (e.g., cells, tissue structures) that should be detected by the algorithm solely by using the provided images without any additional information. The number of markups needed differs from project to project. In case of the osteoclast detection algorithm, about 100 FOVs were manually marked up for osteoclasts. It is essential to draw these markups on a second "layer" (like a "transparency film" on top of the original image) of the target image so that no information from the original image is lost. Alternatively, markups can be stored as simple polygons using popular tools such as Fiji (Schindelin et al. 2015). Either way, manual markups can then be automatically compared to automatically generated masks by the algorithm, ensuring that the newly developed tool and the tissue experts produce comparable results. Supervised machine-learning classifiers like support vector machines, logistic regression, and random forests require ground-truth data as an input to build their decision model.

Before starting to develop an algorithm, the agreement between different human experts has to be confirmed. If this is very low, meaning that there is no consensus between the different human experts, even the best detection algorithm cannot succeed, and therefore development should be postponed, until an agreement between human experts has been achieved. The quality of the ground-truth data can be increased if several human experts provide markups of the same set of images. It is very important that the markups done by humans are performed independently by other experts and by the algorithm developer so that a reliable ground truth can be obtained. A comparison between the experts can be calculated by computing the correlation between the markups. In some cases, like cell recognition covered in this chapter, pixel-level (meaning that the experts actually would have to draw exactly the same lines around the cells) scoring is not very reasonable; therefore, evaluation on a higher level of abstraction, e.g., object-based evaluation, is preferable. This approach counts the number of detected objects and compares it with those that are found in the ground-truth data. Overlaps between detected objects can be used to compute a rough-place agreement beyond the number of detected objects.

In both cases (pixel-/object-based evaluation), having more than one human expert can generate a more objective ground truth (Srivastava et al. 2013), e.g., by performing a majority voting. As the name suggests, majority voting selects those pixels/objects (e.g., osteoclasts) that are marked up by the majority of the human experts.

The next step is the evaluation of the algorithm output. In various scenarios (e.g., segmentation), algorithms have a vast set of parameters. Optimizing them manually is an impossible task, because it would mean evaluating several hundreds of thousands of images by hand. Therefore *exhaustive parameter-optimization* techniques are applied that compare their output with previously created ground-truth data. This technique runs the algorithm with a large set of possible parameter combinations and returns those that have the highest agreement compared to the human experts' ground-truth data. Care must be taken to prevent overfitting, i.e., choosing seemingly optimal parameters, which are only performing well due to chance and only on the images that were used for parameter optimization. Therefore, not all data should be used for this technique – some data should be held back for a final reevaluation of the best parameter settings. This technique is not new and well studied for machine-learning algorithm evaluations: for example, in cross validation, the input data is separated in equally sized sets, employing one for training and one for testing the learned model.

One question that arises is how to compute a score of agreement between the human experts and the algorithm. On pixel level we can calculate the F-score to rank different algorithm output-masks. The F-score (Chinchor and Sundheim 1993) is composed of *precision* and *recall.* Informally, precision represents the percentage of relevant pixels retrieved and is also known as positive prediction rate. Recall on the other hand is also called sensitivity or true positive rate and corresponds to the ratio of correctly classified pixels versus all possible pixels. Formally they are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

TP represents the true positive pixel, those that were marked up by both the expert and the algorithm, whereas TN (true negative) stands for pixels that were not marked by either of them. If the pixel was only detected by the algorithm, then it is called a false positive (FP). Similarly, pixels that were not assigned as belonging to the object of interest by the algorithm but were marked up by the expert are called false negative (FN). With these measures, the *balanced F-score* is defined as:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recal}}{\text{Precision} + \text{Recall}}$$

Other common metrics for pixel-level evaluation of segmentation results are the *Jaccard index* (1912) and *Dice coefficient* (1945). Former is defined as the size of the intersection divided by the size of the union of both sets:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

J denotes the *Jaccard index*. *A* denotes the set of pixels marked up by the expert and *B* the set of pixels detected by the automated segmentation algorithm.

The *Dice coefficient* is a similarity measure closely related to the *Jaccard index* and often used in ecology. It is less prone to outliers and better suited for heterogeneous data sets (McCune et al. 2002). Formally it is defined as follows:

$$s = \frac{2|A \cap B|}{|A| + |B|}$$

where *s* denotes the *Dice coefficient*, *A* the sets of pixels marked up by the expert, and *B* the automated segmentation algorithm.

The above measures were used in recent state-of-the-art literature to evaluate the performance between expert and machine or between various machine-derived segmentations (Hunter et al. 2013; Esteves et al. 2013; Maska et al. 2014). Besides pixel- and object-based comparison, there are other methods, which may be more suitable for specific scenarios. An extensive survey of segmentation methods was published in Zhang (1996). Segmentation is a well-studied topic in computer vision but still a challenging task. Several evaluation methods have been published (Zhang 2001; Smochina 2010; Benes and Zitova 2015). Recently, even crowdsourcing was employed to evaluate segmentation results of experts with various levels in expertise and automated image processing systems (Irshad et al. 2015).

When applying these discussed points to the example of osteoclast detection, difficulties with their practical implementation become obvious. An example markup of a target cell can be seen in Fig. 2.1 (the green line indicates the manually drawn perimeter of an osteoclast). Since the number of osteoclasts is the desired output of the algorithm, an object-based evaluation is more suitable than a pixel-based one. In this case it is less important whether a specific pixel belongs to the osteoclast or to the background as long as the number, general location, and size of the objects (= osteoclasts) agree with the ground-truth data. As mentioned above, to obtain a general detection algorithm which does not only model the specific knowledge of one expert alone, multiple experts with biological background and experience with osteoclast cultures should deliver ground-truth data. In our setting, two experts provided osteoclast markups, and only those osteoclasts that had been identified by both human experts were considered to be "real." When we evaluated the object-level agreement between these two experts in seven different regions (about 70 images) of osteoclast cultures (Fig. 2.3), the mean agreement (matched) between the two experts was $70 \pm 17\,\%$. This shows that even though the manual markup of osteoclasts appears simple in theory (i.e., on an image with a clearly isolated osteoclast as shown in Fig. 2.3), in many other "real-world" images (i.e., in images with osteoclasts and precursor cells in close proximity), the opinions of the human experts as to what qualifies as an osteoclast and what does not can be quite incongruent. Reasons why human-based classification may be error prone were published in Baak (1991) and are discussed later in Sect. 2.5. Comparing our algorithm to the experts' matched markup, 86 % of osteoclasts identified by both human experts were also classified as osteoclasts by the algorithm. In most experiments however, the algorithm classified more osteoclasts than the matched results of

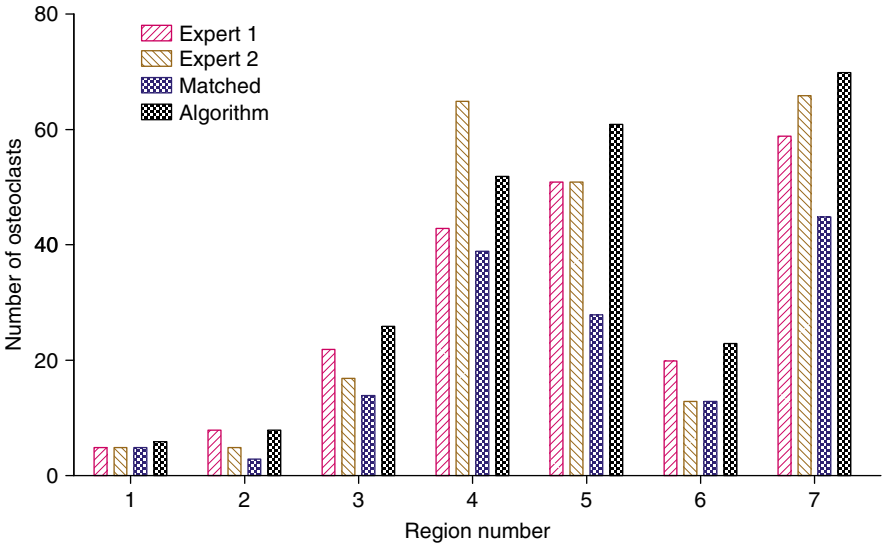

Fig. 2.3 This chart exemplifies the consensus between two human experts. It shows the number of osteoclasts (*y-axis*) detected in seven acquired regions from different cultures (*x-axis*). In each region, the first and the second column represent the osteoclast number detected by each expert (*Expert 1*, *Expert 2*), whereas the third column shows the number of osteoclasts detected congruently by both experts (*matched*)

both human experts (Fig. 2.3). Thus, not only the comparison of the algorithm performance to the matched human experts' ground truth but also the reliability of the ground truth itself will have an impact on the values of precision and recall.

2.4 How to Design an Image-Processing Algorithm Exemplified on Osteoclast Detection in Culture

In this chapter we present the development of an algorithm for osteoclast detection in vitro, starting with the already acquired and stitched images. It follows the general scheme of algorithms for image processing (Fig. 2.4). Most of the existing published image-processing algorithms were derived from this or a similar workflow. Before designing an algorithm, criteria to distinguish between objects of interests (osteoclast) and the remaining objects (precursors) have to be specified. The detailed staining protocol can be found in Sect 2.2. Osteoclasts are defined as multinucleated cells with at least three nuclei (Criterion 1). Additionally, they should not exhibit significant expression levels of F4/80 macrophage marker that identifies only the osteoclast precursor cells (Criterion 2). Such biological criteria should be written down in a document called *customer requirement specification* (CRS) to prevent unexpected results due to misinterpretation by the algorithm developer. The following sections refer to the respective steps of Fig. 2.4: (1) illumination correction, (2) segmentation, (3) post-processing, and (4) labeling).

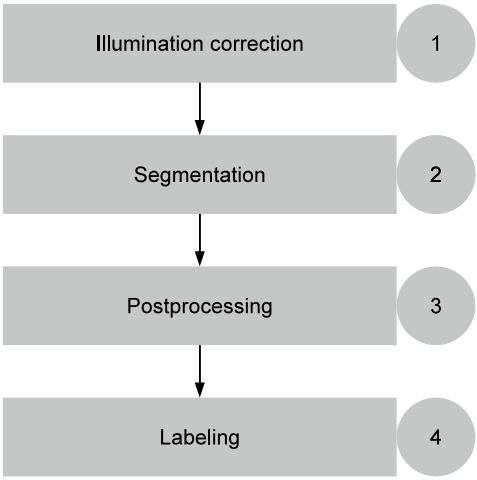

Fig. 2.4 This flowchart demonstrates the general scheme of image-processing algorithms. Four major processing steps must be distinguished

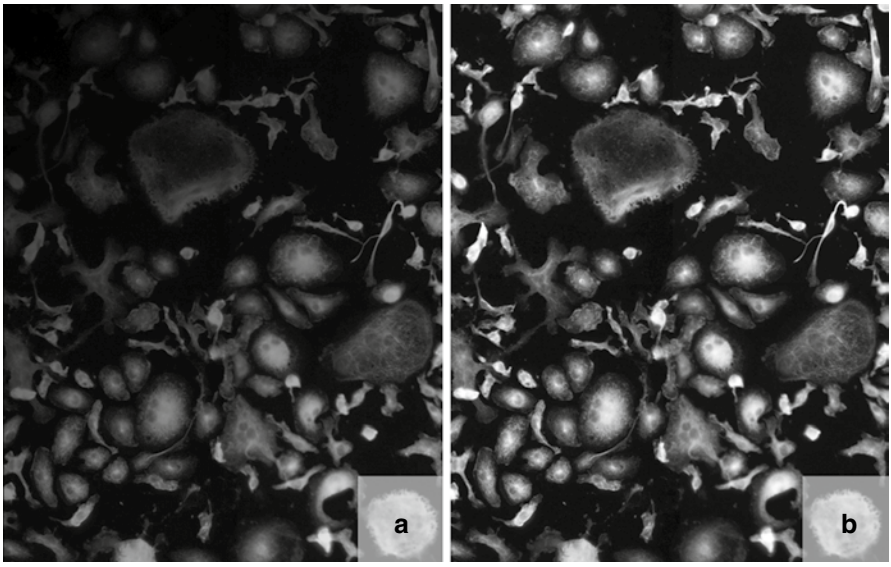

Fig. 2.5 This figure illustrates the effect of a post-acquisition illumination-correction (processing step 1) on images obtained by epifluorescence microscopy. In the original image (**a**), an illumination-gradient is clearly visible from the left upper corner to the right lower corner. This gradient is removed after the application of illumination-correction (**b**)

2.4.1 Illumination Correction

This first step of the detection algorithm is critical, because all further steps will operate on the generated output. Misalignment due to shifts in camera setup or the optical path of the microscope causes uneven illumination (Fig. 2.5a, b) which might not even be visible to the naked eye but can greatly affect the automated

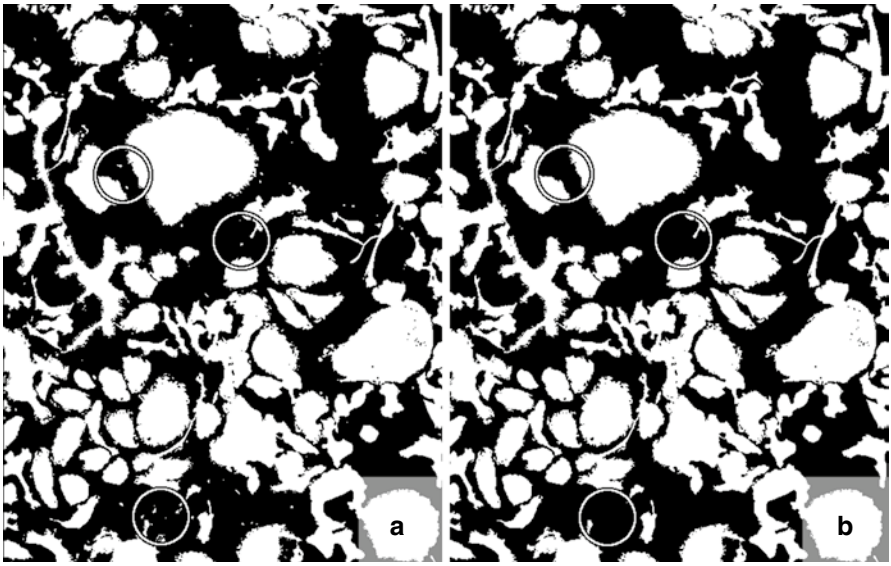

Fig. 2.6 This figure illustrates the results of (**a**) segmentation (processing step 2) and (**b**) post-processing (processing step 3) on images obtained by epifluorescence microscopy. In (**a**), the output image of processing, step 1 (see Fig. 2.5b) has been subjected to automated image segmentation. White structures represent identified objects. This segmentation may introduce artifacts visible as small white blobs that do not belong to actual structures. *Gray circles* indicate example areas where these artifacts are visible. Due to size and shape criteria, they can subsequently be removed resulting in image (**b**)

image analysis. To compensate for this introduced bias, an illumination correction function can be computed (Wu et al. 2008; Wang et al. 2015). It represents a special illumination image that contains the overall pattern of illumination in all future-acquired images. Various publications deal with this process, e.g., (Zhu et al. 2003), Ljosa and Carpenter (2009), and Smith et al. (2015). If the approximation of the illumination function fails, adaptive thresholding can be applied, which is discussed in the next section.

2.4.2 Segmentation

Step 2 splits the pixels in two groups, those belonging to the foreground (cells, including osteoclasts) and those belonging to the background (Fig. 2.6a). This separation can be achieved by applying intelligent thresholding. Examples are the classical triangle algorithm (Zack et al. 1977) or machine-learning-based classifiers with appropriate features (Kapelner et al. 2007). Most of these image segmentation techniques operate on histograms. A histogram is a discrete distribution function of the image's intensity values. It counts the number of gray values that pertain to each of the single categories/bins (intensity values, 0–255 in case of an 8-bit image). The triangle algorithm obtains a threshold to distinguish between background

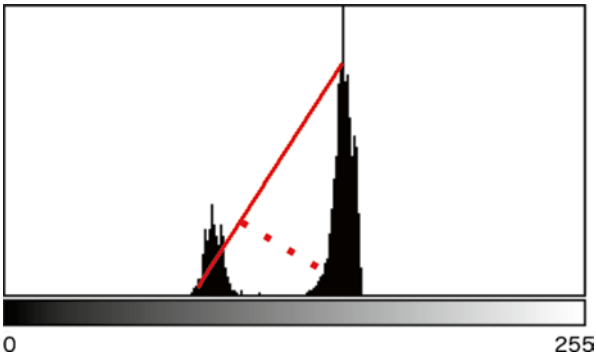

Fig. 2.7 Histogram illustrating the triangle threshold method. The optimal threshold is selected where the maximum distance (*dotted line*) intersects with the base line at the bottom

(< threshold) and foreground (≥ threshold) by computing two local maxima of this histogram. Now these two peaks are connected, and the maximum distance (an orthogonal vector to the line connecting the two peaks) between the line and the histogram is computed (see Fig. 2.7 for an example). The intensity value indicated by the maximum distance is the threshold value. Everything greater or equal (brighter) is considered as foreground. Every value beneath this threshold (darker) is classified as background. The machine-learning-based methods are more complex and require extensive knowledge in learning theory. The interested reader should refer to the publication mentioned above.

If illumination correction (step 1) did not produce an acceptable result, adaptive thresholding can be used to partition images into foreground and background as well. Compared to a single threshold like in the triangle approach, this is a more sophisticated process that chooses different thresholds for each pixel of the image. Because of this method of operation, this is sometimes also called local or dynamic thresholding (Gonzalez and Woods 2008; Stockman and Shapiro 2001; Burger and Burge 2016; Korzynska et al. 2013).

For our osteoclast detection algorithm, we used adaptive thresholding (Liu et al. 2002) and computed local thresholds for each subregion of the image. The threshold is chosen by examining the intensity values of the local neighborhood of each pixel, computing the median intensity. An important parameter of this approach is the neighborhood size which can be determined by thresholding the image with different neighborhood sizes (e.g., 100 pixel, 200 pixel, etc.) and comparing the segmented image with the previously discussed ground-truth markup of the expert. The neighborhood size with the highest concordance is chosen and every subsequent image of the same type processed with the optimized parameter.

2.4.3 Post-processing

Step 3 implements a cleanup step, which removes artifacts (Fig. 2.6a, b) and unwanted cells such as osteoclast precursor cells. This is often achieved by binary operations such as area opening or the computation of morphological features that

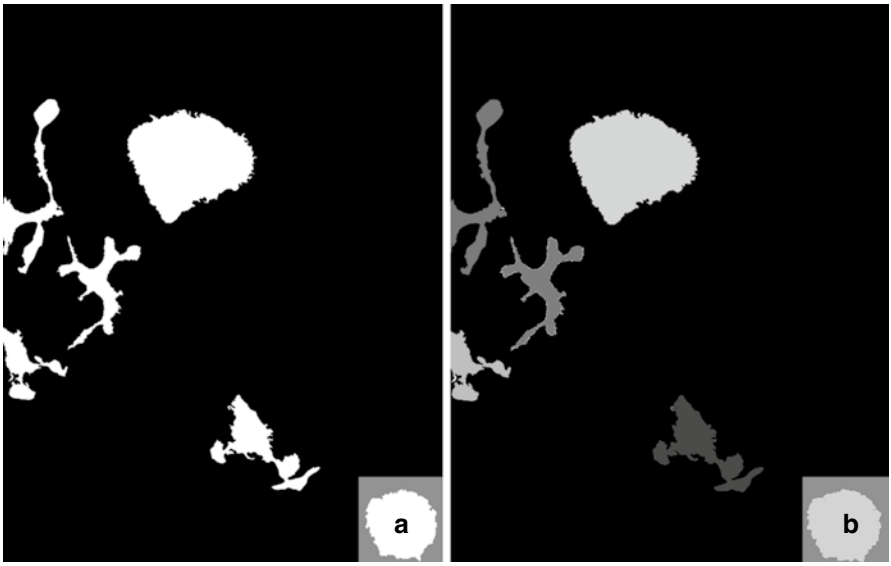

Fig. 2.8 This figure shows the labeling (processing step 4) of images obtained by epifluorescence microscopy. In image (**a**), an enlarged area of the output image of processing step 3 (see Fig. 2.6b) is shown. In image (**b**), the labeling of the segmented image is illustrated. For demonstration purpose, each object is labeled in a different shade of gray. Internally, this would be represented by assigning a unique number to each object

can be used to distinguish between the object of interest and other objects. Area opening removes all objects with an area smaller than a chosen threshold. If this condition is not specific enough, morphological features such as eccentricity, solidity, convex hull, etc., can be used to distinguish between the desired cell and unwanted artifacts. In case of our developed osteoclast detection algorithm, we have also used features from specific staining, i.e., Criterion 1 (osteoclasts: $\geq$3 nuclei) and Criterion 2 (osteoclast: F4/80 mean intensity/cell <20).

2.4.4 Labeling

Step 4 performs a labeling of the remaining cells (Fig. 2.8a, b) (Samet and Tamminen 1988; He 2012). This is done to directly denominate each single cell on the image so that further measurements can be computed by directly addressing the single cells.

As a side note, the image context is also of great importance during feature calculation. Figure 2.9a, c show two osteoclasts. Reasoning from the example at the top, a feature such as the distance between the nuclei may perfectly identify the target cell. Unfortunately, in the same region, another osteoclast can be found (Fig. 2.9c) where the distance is much larger than in the upper example. So if one were to try and detect osteoclasts only in the DAPI channel, i.e., by looking at the nuclei alone, this might produce false positives or negatives. This is why we

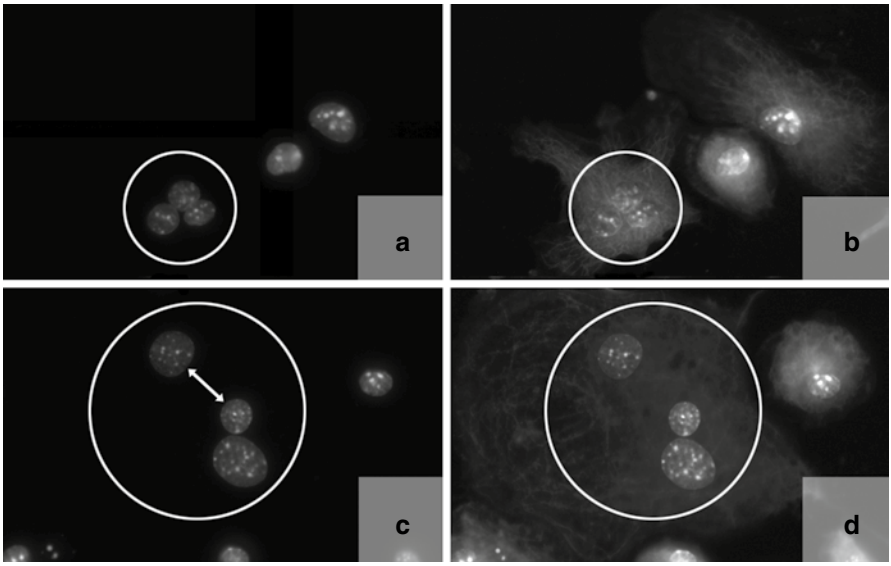

Fig. 2.9 The importance of context is illustrated in this figure. Figures (**a**–**d**) show two different image details of immunofluorescence-labeled osteoclast cultures. *White circles* indicate osteoclasts. During the staining process, cell nuclei are stained with DAPI (**a**, **c**). Additionally, microtubules and a membrane receptor are immunofluorescence-labeled to visualize cell bodies (**b**, **d**). An important criterion of mature osteoclasts is to have three or more nuclei. An algorithm that operates only on the nuclei (DAPI channel) would probably miss the osteoclast in (**c**) due to the larger distance between the nuclei in contrast to (**a**). Taking the additional staining of cell bodies and borders into account, the same nuclei in (**b**) and (**d**) are now allocated to one cell, and consequently these cells are identified as osteoclasts

introduced the additional immunofluorescence stainings of the microtubules, the calcitonin receptor, and the F4/80 macrophage marker to assign the detected nuclei to whole cells (Fig. 2.9b, d).

Having computed all these features, it is now possible to create statistics and draw conclusions about treatment- or disease-related morphological intensity-based difference or staining intensity-based (i.e., associated protein levels) difference. Our final developed algorithm for image-based osteoclast detection has since been incorporated into the StrataQuest image analysis software package.

2.4.5 Practical Example: Influence of Melatonin on Osteoclast Formation

To exemplify the use of this novel automated system for relevant questions in basic biological bone research, we have investigated the effect of pharmacological (μM) doses of melatonin on osteoclast formation in vitro (Bubenik et al. 1998). Using the automated osteoclast detection algorithm in StrataQuest, a multitude of parameters were measured and computed including (mean) intensity values of labeled proteins

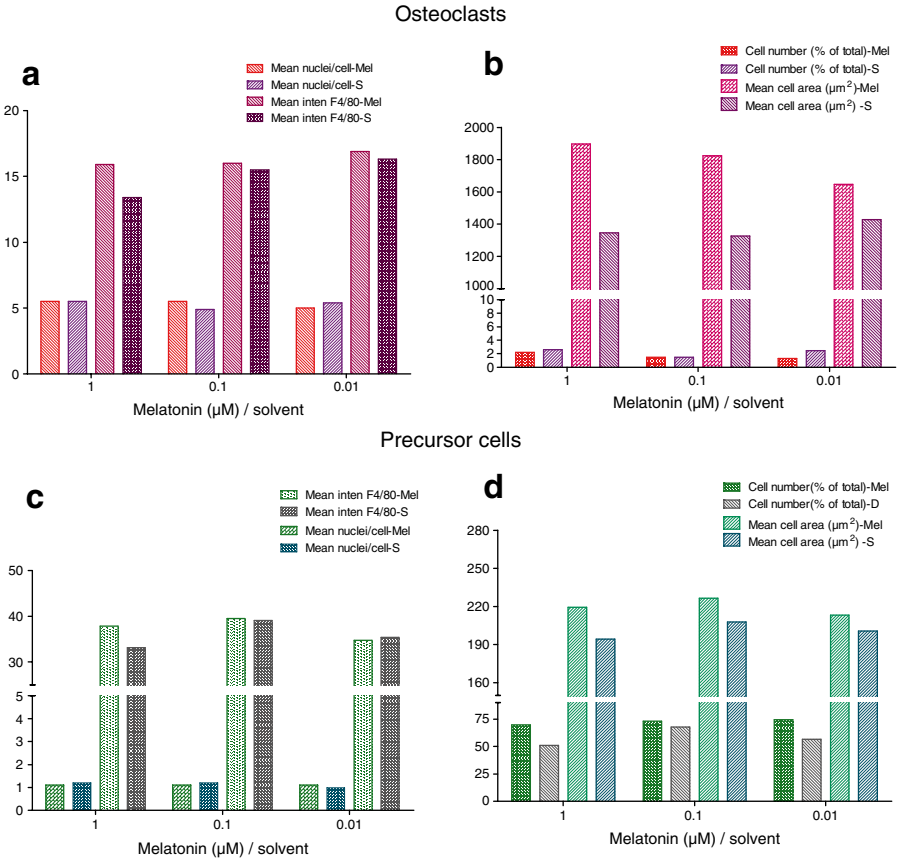

Fig. 2.10 Effect of in melatonin treatment (Mel: 1 μM, 0.1 μM, 0.01 μM) or solvent (S) on cultured osteoclasts (**a**, **b**) and precursor cells (**c**, **d**). (**a**, **c**) Mean number of nuclei/cell and mean intensity levels of F4/80 staining/cell. (**b**, **d**) Relative (% of total) cell number and mean cell area (μm^2). Data from a typical experiment for multinucleated cells with mean intensity levels of F4/80 staining <20 (1.3–2.6 % of cells) and mononucleated cells (precursor cells) with mean intensity levels of F4/80 staining >20 (51–74 % of cells) are shown

per cell (exemplified for F4/80 staining, Fig. 2.10a, c), number of nuclei per cell (Fig. 2.10a, c), total or relative number of individual cells, or mean cell area (Fig. 2.10b, d). Results are displayed for two out of the four discriminated cell populations, namely, the multinucleated osteoclasts with mean F4/80 intensity levels <20 (arbitrary threshold, range 0–255) and the mononucleated precursor cells with mean F4/80 intensity levels >20 (Fig. 2.10 upper and lower graphs, respectively). While, in analogy to published data (Koyama et al. 2002), we could not see any influence of melatonin on osteoclast number (Fig. 2.10b); we found a strong increase ranging from 130 % of control values with 10 nM melatonin to 200 % of control values with 1 μM melatonin in the mean area of osteoclasts after treatment with melatonin, an effect which was not seen for the precursor cells (Fig. 2.10b). In

correlation, also the mean number of nuclei/osteoclast increased up to 140 % of control values (Fig. 2.10a). The classical evaluation by humans would have only retrieved the number of osteoclasts/area or well, which in our experiment has not been influenced. The observation that the mean area of and the mean number of nuclei/osteoclast increased (but not of the precursor cells) indicates that melatonin has a stimulating effect on the fusion of osteoclast precursor cells to mature osteoclasts. Whether the morphological alterations are associated with an increased bone-resorbing activity of the osteoclasts needs to be determined in functional assays, e.g., measurement of pit formation on dentine slices (Takahashi et al. 2007). Nevertheless, the results gained by automated evaluation for the first time demonstrate a direct effect of melatonin on bone-resorbing cells, which together with the well-established stimulatory influence on osteoblast activity would greatly influence bone turnover.

2.5 Common Pitfalls

Although high-end microscopy technologies open up new ways to examine tissue and cell culture samples, it also requires detailed knowledge of biology, optics, electronics, and computer science. This includes appropriate sampling of the tissues, optimized cell-culture conditions, furthermore well-chosen fixation, and staining protocols. It also includes the selection of the most appropriate microscopic equipment for acquisition of the specific experiment (Pearson 2007). However, in this section we want to focus on those common problems that occur during image acquisition and evaluation. We discuss pitfalls like uneven illumination, uneven staining, touching clumps of nuclei, and psychological aspects resulting in fallacies due to gestalt laws. For basics about microscopy and optics, please refer to Spector and Goldman (2006).

2.5.1 Imaging-Based Errors

2.5.1.1 Illumination

To perform a quantitative measurement, each step of image acquisition has to be as exact as possible. Noise is often a result of a misaligned light source and increases the error rate in all following algorithm steps. Therefore, having an evenly illuminated image is of great importance. To achieve this goal, calibration slides can be used to align the light source properly before recording microscopy images. Applying an illumination-correction function (as discussed in Sect. 4.1) afterwards is a measure of last resort since it modifies the intensity values of the image and might distort the "real" microscopic picture. If this is done in the channel containing the target protein, the researcher has to consider that she/he may have created an artificial (= false) staining due to the applied correction function. Using these results for further statistical analysis is problematic and may lead to wrong conclusions. The most common optical problem, nonaligned condensers, causes a type of uneven

illumination that is almost impossible to repair with image-processing techniques. A priori detection and elimination of illumination gradients before recording images therefore save a lot of effort and avoid unnecessary image manipulation, thus increasing the significance of derived quantitative biological results.

2.5.1.2 Acquisition Parameters

Acquiring the same sample on different instruments or with different settings on the same instrument often results in different-looking images. Depending on the settings and the experience of the user who controls the microscope, image quality may vary. Currently, there is no standard (Yagi and Gilbertson 2005), so the comparability between various acquisition systems is impossible. An example simulating the effect of different acquisition settings on the same FOV is shown in Fig. 2.11. The first image (Fig. 2.11a) of this figure is in focus and has balanced color values as it should be. Figure 2.11b is out of focus; consequently segmentation will prove difficult and unreliable because borders are blurred. To prevent out-of-focus images, techniques like the so-called extended focus (Abrahamsson et al. 2006) can be used: first, a stack of images of the same FOV on five different focal planes is acquired. A subsequent fusing step merges all images per stack and takes only those parts that are in focus on each plane. As a result, the amount of unfocused cells is reduced to a minimum. Figure 2.11c illustrates overexposure. Saturated areas are visible as

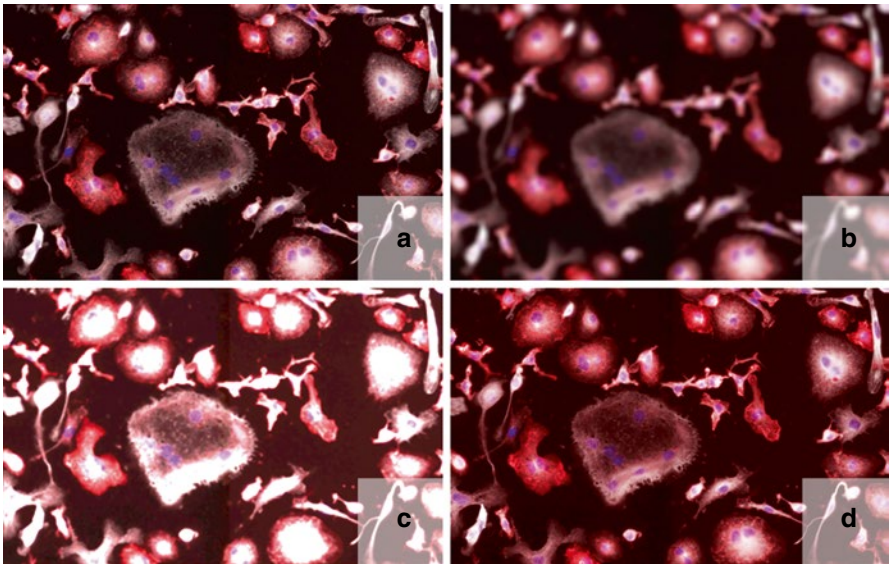

Fig. 2.11 This figure illustrates the effect of acquisition conditions in image processing. Image (**a**) represents an ideal acquisition, showing the cells in focus with balanced color values. In image (**b**), cells are out of focus, and their borders are no longer clearly visible. Image (**c**) is overexposed (exposure time set too long). Saturated areas make further quantitative analysis impossible. Finally, image (**d**) has a significantly increased red value (background) that may mislead the observer. Note: These images have been processed for demonstration purposes

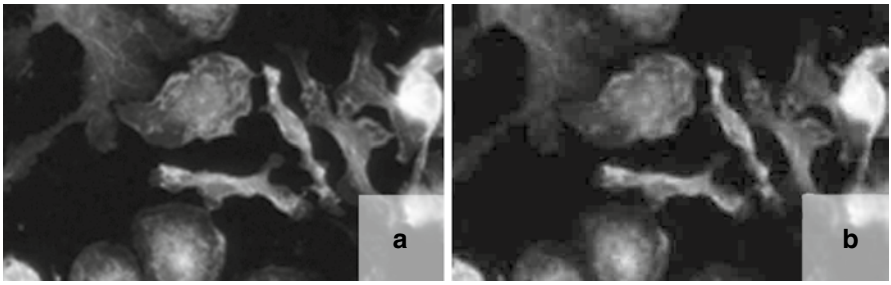

Fig. 2.12 The importance of file formats in image processing is illustrated in this figure. Images (**a**, **b**) represent the same image detail derived from an immunofluorescence-labeled osteoclast culture. While image (**a**) was stored as PNG (with lossless compression), image (**b**) was saved as JPEG. JPEG may reduce image quality dramatically and may introduce rectangular artifacts that could affect machine-learning classification

white spots and prevent useful feature computation. An increased red level is illustrated in Fig. 2.11d. This may mislead the observer to draw incorrect conclusions about the expression of a certain marker.

2.5.1.3 File Formats

Another pitfall can be the file format used to store the image. Lossless formats should always be preferred to prevent quantification artifacts. A good choice for storing high-quality images would be the portable network graphics (PNG) (Fig. 2.12a). This format can be read by a variety of tools and supports lossless compression. Besides PNG, the tagged image file format (TIFF) is popular for storing acquired images. Compared to PNG, it offers a wide range of storage options, which is also the drawback of this format. It cannot be guaranteed that other applications which offer TIFF support are able to read and correctly interpret the chosen TIFF settings. The most commonly used Joint Photographic Experts Group (JPEG) graphics format has to be avoided (Fig. 2.12b). The intention of its compression is to remove details that are not visible to the human eye. Obviously, this alters intensity values and can limit the applicability of machine-learning and image-processing techniques as well as quantification of biological features.

For the osteoclast detection, we selected PNG due to the portability of the format to different operating systems like Windows, Mac OS X, and Linux, since preliminary experiments indicated about 30 % less disk space usage than when using compressed TIFF images.

2.5.2 Errors Related to the Gestalt Laws

The origin of the term Gestalt (the essence of an entity's complex form) goes back to Ernst Mach in 1886 (Mach 1886). Since that time, gestalt laws are extensively examined in psychology. An example of these laws is grouping of objects due to proximity, similarity, closure, good continuation, and connectedness. A summary

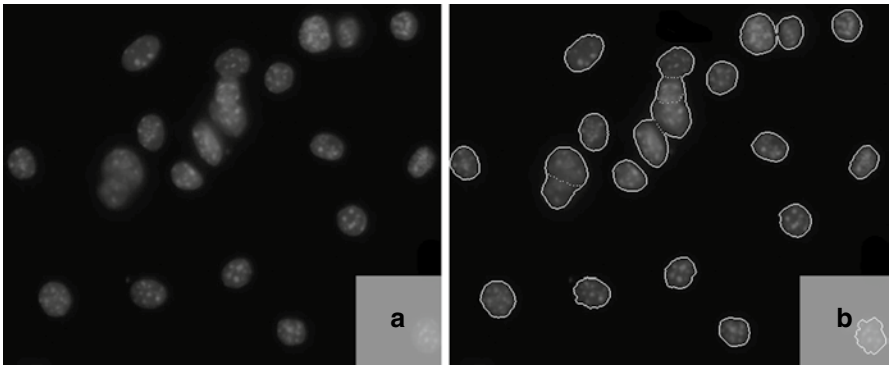

Fig. 2.13 This figure illustrates the limitations of automated image segmentation in comparison to human-object recognition. Figure (**a**) shows an image detail of an osteoclast culture, where cell nuclei are labeled with DAPI. *Clusters* of nuclei, i.e., overlapping nuclei, are visible. The result of a segmentation of these nuclei is indicated in image (**b**). The detected perimeters derived from automated image segmentation are drawn as *solid white lines*, while the *dotted lines* represent the additional separation of the overlapping nuclei, as a human would intuitionally draw them

published in Scientific American describes and illustrates these gestalt laws comprehensively (Rock and Palmer 1990). Especially when ground-truth data is created, gestalt laws play an important role. Human recognition of shapes is still a research field with many open questions. One way to reduce the effect of gestalt laws is to make people aware of it: Training them with examples can increase the output quality of ground-truth data, thus improving detection accuracy of the algorithm. As a result, interdisciplinary projects should be preferred to avoid such imprinted pitfalls.

Up to this day, there are scenarios where image processing cannot compete with human intuition. One frequently occurring example in osteoclast detection is the problem of touching clumps of nuclei (Moffat et al. 2006). Figure 2.13a exemplifies a case where image processing may fail, although nuclei can be intuitively separated by a human (Fig. 2.13b), and different human observers are remarkably consistent in where they separate the given nuclei. Analyses that require identification of a single nucleus fail if no proper segmentation of these clumps is available. There are several approaches to divide clumps in a single nucleus in culture and tissues (Rogojanu et al. 2010; Zhang et al. 2015; Sheeba et al. 2014) like applying watershed algorithms (Sheeba et al. 2014), cutting of nuclei due to angles in the morphological shape (Cloppet and Boucher 2010; Wang et al. 2012), level-set-based processes (Xiong et al. 2006), or dynamic programming that tries to model the human expertise (Nandy et al. 2007). However, each of these approaches has scenarios where they fail, so currently there is no computational algorithm available that can handle all different cases of clumps.

Trying to model this method is tricky due the fact that it is not yet known how the human brain recognizes objects (Liu et al. 2009). Figure 2.14 illustrates an idealized decision-making process of a human. In real life, this process is assumed to be less

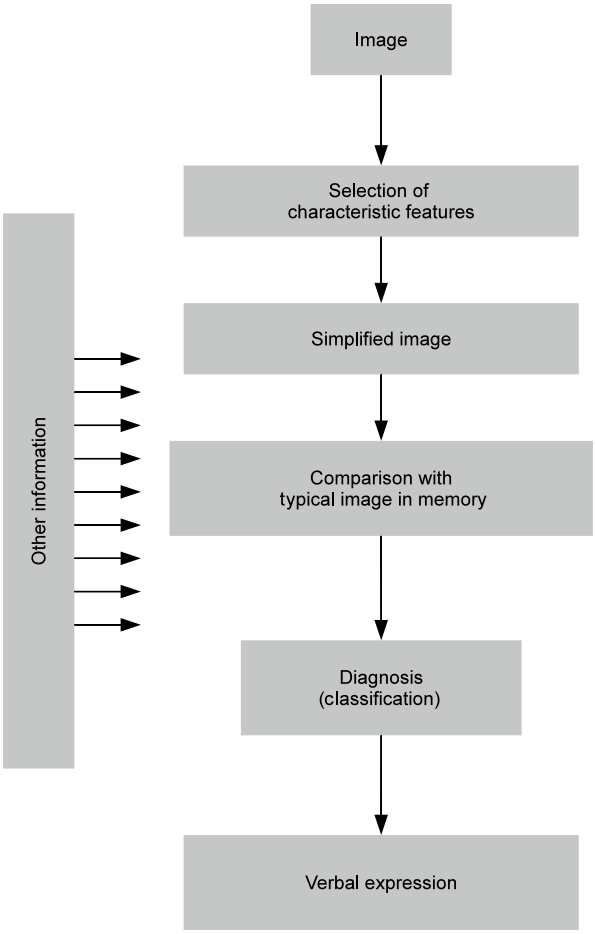

Fig. 2.14 This flowchart shows an assumed human diagnostic process of an idealized decision-making situation (Modified after Baak (1991))

structured and contains trained template recognition, which is believed to be present in the human subconscious and instantly available (Baak 1991; Lennert and Stein 1981) – an example would be recognition of numbers or letters in literate humans. In contrast to the human brain, slight variations in size, staining, orientation, and illumination are not accounted for by the computer and therefore result in poor recognition performance. Currently, image processing uses features (e.g., curvature, intensity changes, homogenous textures, edges, etc.) that seem to be too different and too "weak" to model the human performance of perception.

Nevertheless, visual perception is not the only source of error; verbal expression differs from expert to expert. An example would be the size of osteoclasts. One human expert may assume that a "huge" osteoclast is one with up to 6 nuclei, whereas for another expert, "huge" osteoclast means more than 16 nuclei. Obviously,

development of an algorithm requires background knowledge of the target cell structure and texture. Ideally, the computer scientist obtains this information during interviews with the biological expert, and they create a CRS together at the beginning of the project instead of relying on verbal-communicated information only. Quantitative evaluation is given by ground-truth data which should agree with explicit knowledge given in the CRS, and over the course of the project, the CRS should be regularly discussed and possibly adapted. The interpretation of the given features to classify the target cells varies depending on the final knowledge of the algorithm engineer. This was shown for pathologists in Livesey et al. (1978), Pool et al. (1979). However, nowadays when pathological interpretations are quantified by computer-based evaluation, computer scientists take the place of the pathologists by developing tools to support their diagnoses. Clearly, they face the same problems with less medical and biological experience, and additionally they interpret the data based on their technical background which may lead to problems of overgeneralization.

2.5.3 Benefits of Automated Osteoclast Detection

Despite these pitfalls, automated segmentation and analysis of osteoclasts as exemplified in this chapter has significant advantages. Large amount of data is processable – for the results shown in Fig. 2.10, we have processed regions up to 42 mm^2, and the only limits for further scale-up are the acquisition of the images and the computer's processing and memory capacity. Our computer-based evaluation is faster by about two orders of magnitude compared to a trained human expert. Additionally, the result after manual quantification is just the number of osteoclasts, whereas our algorithm yields many more informative measures, such as total cell number, total and relative numbers of osteoclasts and their precursor cells, total area of all cells, total and relative area of osteoclasts and their precursor cells, numbers of nuclei, quantification of associated proteins, and many morphological and statistical features. Furthermore, it is highly unlikely that the human quantification on large regions is reproducible or consistent if compared to another human expert (due to fatigue and different interpretations by different humans). However, reapplying an algorithm to the same set of images always yields the identical result.

Conclusion

Applying image processing and machine-learning techniques to biological and medical images can improve the quality of research and diagnostics dramatically; automated analysis produces consistent quantitative measures. Small differences not visible for the human eye, but possibly linked to disease states, can be detected. Currently applied visual inspection normally produces an overall score rather than measuring each cell. The human mind also cannot keep track of the multiple informative measures of cells or tissue and is generally less able to integrate many weak predictive measures. It should also be emphasized again that machine-based analysis is more efficient after development and can operate 24 h, 7 days a week.

One domain that especially benefits from such systems is bone research. Osteoclast quantification is currently done manually, so that large-scale experiments cannot be conducted. The intra- and inter-variability between experts is normally very high, which is in contrast to an automated system that detects and quantifies these cells, improves the quality of the result, and is always consistent. Applying the algorithm, in addition, enables to measure parameters, which cannot be assessed manually. Indeed, practical application of the novel osteoclast detection system in the software StrataQuest enabled us to identify a direct influence of the indolamine melatonin on in vitro murine osteoclast size and multinuclearity which would not have been accessible without in silico-based image analysis.

By extension, together with automated staining systems, a reproducible, validated, and fully automatic workflow for medium to high throughput evaluation of basic osteoclast biology, but also routine clinical screening, is conceivable. Of course, legal issues have to be considered before such algorithms can successfully be applied to standard diagnosis in hospitals, and the results of automated quantifications are dependent on a number of parameters like correct staining and acquisition. Image processing will never replace human experts completely, because the final diagnosis and interpretation is still up to human expertise, but it can relieve the scientists or physicians from a huge amount of repetitive work and at the same time increase the significance of the obtainable results.

Acknowledgements This project was funded by FFG (Bridge 818094). We also want to express our deepest thanks to Caroline Brünner-Kubarth, Kathrin Burger, Andrea Nußbaumer, Ayse Okay, Magdalena Pilz, Radu Rogojanu, and Roland Stumberger for their contributions and dedication to our project.

References

Abrahamsson, S., Usawa, S., & Gustafsson, M. (2006) A new approach to extended focus for high-speed, high-resolution biological microscopy. In Biomedical Optics 2006 (pp. 60900N-60900N). International Society for Optics and Photonics.

Akatsu T, Tamura T, Takahashi N, Udagawa N, Tanaka S, Sasaki T, Yamaguchi A, Nagata N, Suda T (1992) Preparation and characterization of a mouse osteoclast-like multinucleated cell population. J Bone Miner Res 7(11):1297–1306

Andersson GN, Marks SC Jr (1989) Tartrate-resistant acid ATPase as a cytochemical marker for osteoclasts. J Histochem Cytochem 37(1):115–117

Baak JPA (1991) Manual of quantitative pathology in cancer diagnosis and prognosis. Springer, Berlin/New York, [1991] ©1991

Bakushinskiy A, Goncharsky A (2012) Ill-posed problems: theory and applications. Springer, Netherlands

Benes M, Zitova B (2015) Performance evaluation of image segmentation algorithms on microscopic image data. J Microsc 257(1):65–85

Bubenik GA, Blask DE, Brown GM, Maestroni GJ, Pang SF, Reiter RJ, Viswanathan M, Zisapel N (1998) Prospects of the clinical utilization of melatonin. Biol Signals Recept 7(4):195–219

Burger W, Burge MJ (2008) Digital image processing – an algorithmic introduction using java. Springer, London, XX, 565 pp

Burger W, Burge MJ (2016) Digital image processing: an algorithmic introduction using java. Springer, London

Chinchor N, Sundheim B (1993) MUC-5 evaluation metrics. Association for Computational Linguistics, Baltimore, pp 69–78

Cloppet F, Boucher A (2010) Segmentation of complex nucleus configurations in biological images. Pattern Recogn Lett 31(8):755–761

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26(3):297–302

Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, Murphy RF, Peng H, Plant AL, Roysam B, Stuurman N, Swedlow JR, Tomancak P, Carpenter AE (2012) Biological imaging software tools. Nat Methods 9(7):697–710

Esteves T, Oliveira MJ, & Quelhas P (2013) Cancer cell detection and morphology analysis based on local interest point detectors. In Iberian Conference on Pattern Recognition and Image Analysis. Springer Berlin Heidelberg, 624–631

Gonzalez RC, Woods RE (2008) Digital image processing. Prentice Hall, Upper Saddle River

He L, Chao Y, & Suzuki K. (2012) A new two-scan algorithm for labeling connected components in binary images. In Proceedings of the World Congress on Engineering, 2:1141–1146

Holmes S, Kapelner A, Lee PP (2009) An interactive java statistical image segmentation system: GemIdent. J Stat Softw 30(10):i10

Hunter LA, Krafft S, Stingo F, Choi H, Martel MK, Kry SF, Court LE (2013) High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images. Med Phys 40(12):121916

Irshad H, Montaser-Kouhsari L, Waltz G, Bucur O, Nowak JA, Dong F, Knoblauch NW, Beck AH (2015) Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. Pac Symp Biocomput: (p. 294). NIH Public Access.

Jaccard P (1912) The distribution of the flora in the alpine zone. 1. New Phytol 11(2):37–50

Zhu J, Liu B, & Schwartz SC (2003) General illumination correction and its application to face normalization. In Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on. IEEE (Vol. 3, pp. III–133)

Kapelner A, Lee PP, & Holmes S (2007) An interactive statistical image segmentation and visualization system. In Medical Information Visualisation-BioMedical Visualisation, 2007. MediVis 2007. International Conference on. IEEE 81–86

Korzynska A, Roszkowiak L, Lopez C, Bosch R, Witkowski L, Lejeune M (2013) Validation of various adaptive threshold methods of segmentation applied to follicular lymphoma digital images stained with 3,3′-Diaminobenzidine&Haematoxylin. Diagn Pathol 8:48

Koyama H, Nakade O, Takada Y, Kaku T, Lau KH (2002) Melatonin at pharmacologic doses increases bone mass by suppressing resorption through down-regulation of the RANKL-mediated osteoclast formation and activation. J Bone Miner Res 17(7):1219–1229

Krig S (2014) Ground truth data, content, metrics, and analysis. Apress, Berkeley, pp 283–311

Lennert K, Stein H (1981) Histopathology of non-Hodgkin's lymphomas: based on the kiel classification. Springer, Berlin

Liu F, Song X, Luo Y, Hu D (2002) Adaptive thresholding based on variational background. Electron Lett 38(18):1017–1018

Liu H, Agam Y, Madsen JR, Kreiman G (2009) Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62(2):281–290

Livesey AE, Sutherland FI, Brown RA, Beck JS, Macgillivray JB, Slidders W (1978) Cytological basis of histological typing of diffuse Hodgkin's disease. Demonstration of an implied misnomer in the terminology of the Rye classification. J Clin Pathol 31(6):551–559

Ljosa V, Carpenter AE (2009) Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening. PLoS Comput Biol 5(12):e1000603

Mach E (1886) Beiträge zur Analyse der Empfindungen. G. Fischer, Jena

Maria S, Witt-Enderby PA (2014) Melatonin effects on bone: potential use for the prevention and treatment for osteopenia, osteoporosis, and periodontal disease and for use in bone-grafting procedures. J Pineal Res 56(2):115–125

Marino S, Logan JG, Mellis D, Capulli M (2014) Generation and culture of osteoclasts. Bonekey Rep 10(3):570

Martin DR, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. EECS Department, University of California, Berkeley

Maska M, Ulman V, Svoboda D, Matula P, Matula P, Ederra C, Urbiola A, Espana T, Venkatesan S, Balak DM, Karas P, Bolckova T, Streitova M, Carthel C, Coraluppi S, Harder N, Rohr K, Magnusson KE, Jalden J, Blau HM, Dzyubachyk O, Krizek P, Hagen GM, Pastor-Escuredo D, Jimenez-Carretero D, Ledesma-Carbayo MJ, Munoz-Barrutia A, Meijering E, Kozubek M, Ortiz-de-Solorzano C (2014) A benchmark for comparison of cell tracking algorithms. Bioinformatics 30(11):1609–1617

McCune B, Grace JB, Urban DL (2002) Analysis of ecological communities. MjM Software Design, Gleneden Beach

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, Carpenter AE, Foo SY, Stewart SA, Stockwell BR, Hacohen N, Hahn WC, Lander ES, Sabatini DM, Root DE (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124(6):1283–1298

Nandy K, Gudla PR, & Lockett SJ (2007) Automatic segmentation of cell nuclei in 2D using dynamic programming. In Proceedings of 2nd Workshop on Microsopic Image Analysis with Applications in Biology

Ostrowska Z, Kos-Kudla B, Swietochowska E, Marek B, Kajdaniuk D, Ciesielska-Kopacz N (2001) Influence of pinealectomy and long-term melatonin administration on GH-IGF-I axis function in male rats. Neuro Endocrinol Lett 22(4):255–262

Pearson H (2007) The good, the bad and the ugly. Nature 447(7141):138–140

Pool CW, Diegenbach PC, Ockeloen BJ (1979) Quantitative succinate-dehydrogenase histochemistry. II. A comparison between visual and quantitative muscle fibre typing. Histochemistry 64(3):263–272

Radio NM, Doctor JS, Witt-Enderby PA (2006) Melatonin enhances alkaline phosphatase activity in differentiating human adult mesenchymal stem cells grown in osteogenic medium via MT2 melatonin receptors and the MEK/ERK (1/2) signaling cascade. J Pineal Res 40(4):332–342

Rasband WS (1997–2015) ImageJ. U. S. National Institutes of Health, Bethesda

Rock I, Palmer S (1990) The legacy of Gestalt psychology. Sci Am 263(6):84–90

Rogojanu R, Bises G, Smochina C, & Manta, V (2010) Segmentation of cell nuclei within complex configurations in images with colon sections. In Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on. IEEE 243–246

Samet H, Tamminen M (1988) Efficient component labeling of images of arbitrary dimension represented by linear bintrees. IEEE Trans Pattern Anal Mach Intell 10(4):579–586

Schindelin J, Rueden CT, Hiner MC, Eliceiri KW (2015) The ImageJ ecosystem: an open platform for biomedical image analysis. Mol Reprod Dev 82(7-8):518–529

Schmid M, Dufner B, Durk J, Bedal K, Stricker K, Prokoph LA, Koch C, Wege AK, Zirpel H, van Zandbergen G, Ecker R, Boghiu B, Ritter U (2015) An emerging approach for parallel quantification of intracellular protozoan parasites and host cell characterization using TissueFAXS cytometry. PLoS One 10(10):e0139866

Sheeba F, Thamburaj R, Mammen JJ, & Nagar AK (2014) Splitting of Overlapping Cells in Peripheral Blood Smear Images by Concavity Analysis. In International Workshop on Combinatorial Image Analysis. Springer International Publishing 238–249

Smith K, Li Y, Piccinini F, Csucs G, Balazs C, Bevilacqua A, Horvath P (2015) CIDRE: an illumination-correction method for optical microscopy. Nat Methods 12(5):404–406

Smochina C, Manta V, Rogojanu R (2010) New discrepancy measure for evaluation of segmentation quality, In Proc. 11th IASTED International Conference on Computer Graphics and Imaging, Innsbruck, Austria, track 679–053

Solomon C, Breckon T (2011) Fundamentals of digital image processing: a practical approach with examples in matlab. Wiley Publishing, Chichester

Spector DL, Goldman RD (2006) Basic methods in microscopy protocols and concepts from cells: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Srivastava G, Yoder JA, Park J, Kak AC (2013) Using objective ground-truth labels created by multiple annotators for improved video classification: a comparative study. Comput Vis Image Underst 117(10):1384–1399

Stadler M, Walter S, Walzl A, Kramer N, Unger C, Scherzer M, Unterleuthner D, Hengstschlager M, Krupitza G, Dolznig H (2015) Increased complexity in carcinomas: analyzing and modeling the interaction of human cancer cells with their microenvironment. Semin Cancer Biol 35:107–124

Stockman G, Shapiro LG (2001) Computer vision. Prentice Hall PTR, Upper Saddle River

Takahashi N, Udagawa N, Kobayashi Y, Suda T (2007) Generation of osteoclasts in vitro, and assay of osteoclast activity. Methods Mol Med 135(1543-1894 (Print)):285–301

van de Wijngaert FP, Tas MC, Burger EH (1987) Characteristics of osteoclast precursor-like cells grown from mouse bone marrow. Bone Miner 3(2):111–123

Wang H, Zhang H, Ray N (2012) Clump splitting via bottleneck detection and shape classification. Pattern Recogn 45(7):2780–2787

Wang HF, Wang GP, Wang XY, Wang BJ, Zhao XD (2015) Novel background calibration algorithm for image in non-uniform illumination field. Imaging Sci J 63(5):285–289

Wu Q, Merchant F, Castleman K (2008) Microscope image processing. Elsevier Inc., Amsterdam

Xiong G, Zhou X, Ji L, Bradley P, Perrimon N, & Wong S (2006) Segmentation of drosophila RNAi fluorescence images using level sets. In 2006 International Conference on Image Processing. IEEE 73–76

Yagi Y, Gilbertson JR (2005) Digital imaging in pathology: the case for standardization. J Telemed Telecare 11(3):109–116

Zack GW, Rogers WE, Latt SA (1977) Automatic measurement of sister chromatid exchange frequency. J Histochem Cytochem 25(7):741–753

Zhang YJ (1996) A survey on evaluation methods for image segmentation. Pattern Recog 29(8):1335–1346

Zhang YJ (2001) A review of recent evaluation methods for image segmentation. Kuala Lumpur, Malaysia, pp 148–151

Zhang C, Sun C, Su R, Pham TD (2015) Clustered nuclei splitting via curvature information and gray-scale distance transform. J Microsc 259(1):36–52