Image Segmentation, Processing and Analysis in Microscopy and Life Science

Carolina Wählby

Abstract Microscopes have been used for more than 400 years to understand biological and biomedical processes by visual observation. Science is the art of observing, but science also requires measuring, or quantifying, what is observed. Research based on microscopy image data therefore calls for methods for quantitative, unbiased, and reproducible extraction of meaningful measurements describing what is observed. Digital image processing and analysis is based on mathematical models of the information contained in image data, and allows for automated extraction of quantitative measurements. Automated methods are reproducible and, if applied consistently and accurately across experiments with positive as well as negative controls, also unbiased. Digital image processing is further motivated by the development of scanning microscopes and digital cameras that can capture image data in multiple spatial-, time-, and spectral-dimensions, making visual assessment cumbersome or even impossible due to the complexity and size of the collected data.

The process of analyzing a digital image is usually divided into several steps, where the objects of interest are first identified, or 'segmented', followed by extraction of measurements and statistical analysis. This chapter starts from the basics of describing images as matrices of pixel intensities. Emphasis is thereafter put on image segmentation, which is often the most crucial and complicated step. A number of common mathematical models used in digital image processing of microscopy images from biomedical experiments are presented, followed by a brief description of large-scale image-based biomedical screening.

Keywords Image cytometry • Fluorescence microscopy • Cell segmentation • Image analysis

C. Wählby (✉)
Centre for Image Analysis, Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Uppsala, Sweden
e-mail: carolina@cb.uu.se

V. Zazzu et al. (eds.), *Mathematical Models in Biology*,
DOI 10.1007/978-3-319-23497-7_1

1 Pixels and Color Channels

A digital image is not continuous, but consists of discrete picture elements, or pixels. A typical fluorescencemicroscopy image is built up of multiple fluorescence channels, each representing a separate fluorescence stain, usually bound to DNA or an antibody probing a specific protein or subcellular structure. Figure 1 shows a fluorescence microscopy image where cell nuclei are stained with DAPI binding DNA, and red and green dots representing mRNA molecules (for details see [12]). Imagine that the goal of the analysis is to count the number of red and green dots per cell. The color image in Fig. 1 can be split into its constituent image channels, leading to one image representing the red, green and blue fluorescence respectively. If we take a closer look at the red channel, and zoom in on one of the dots, we can see that the image is built up of square picture elements, or 'pixels' for short, see Fig. 2. Each of these pixels is represented as a number in the computer, where a higher number means a brighter pixel, and the whole image can be thought of as a matrix of numbers. In a color image, the three image channels represent the

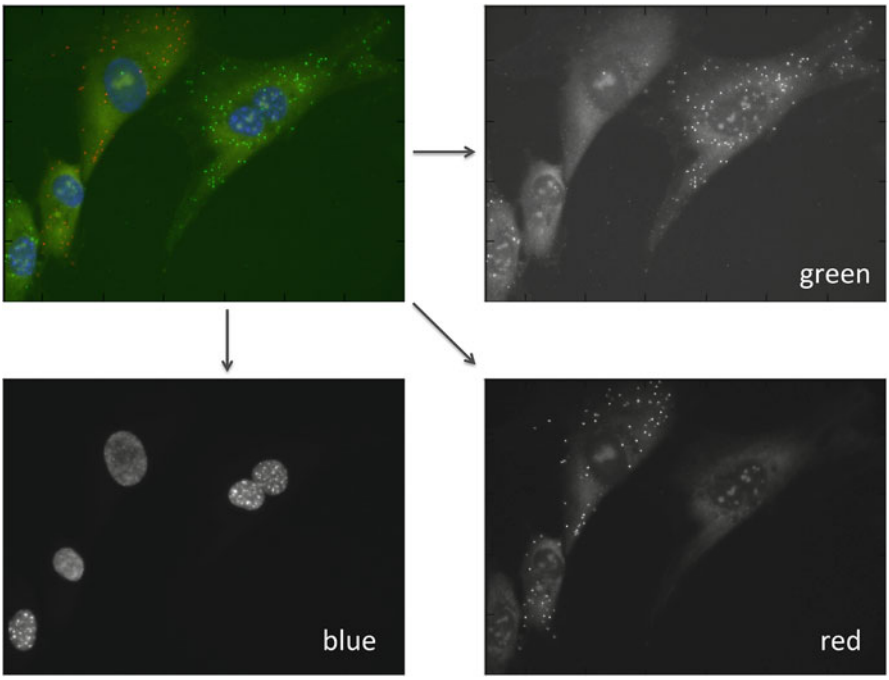

Fig. 1 Using three different filter sets, three different fluorescence labels were imaged using fluorescence microscopy. Top left is a composite image of all three image channels; cell nuclei are stained with DAPI binding DNA, and *red* and *green dots* represent mRNA molecules (for details see [12]). Due to autofluorescence and unspecific fluorophore binding, the cells' cytoplasms can be seen as a weak background staining in the *red* and *green* image channels

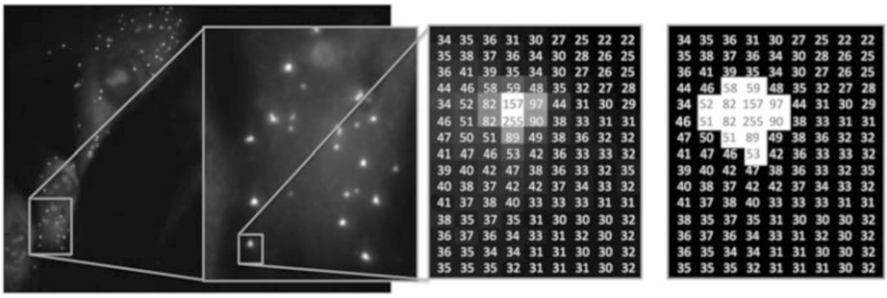

Fig. 2 An image is built up of pixels. If we zoom in on a sub part of the larger image, we see that the image is built up of square pixels, where the graylevel, or intensity of each pixel, can be represented as a number in the computer memory. Image segmentation thresholding (*left*) assigns the maximum value (*white*) to all pixels above a given threshold (in this case intensities >50), while other pixels are assigned the minimum value (*black*)

amount of red, green and blue respectively, and any image analysis operation can work either on a single image channel, or a combination of multiple image channels in two or more spatial dimensions, as well as with time sequences. Here we focus on operations that work on a single channel in two dimensions, but the general idea of images as matrices of pixels, each represented by a number, holds true in any number of dimensions.

2 Image Segmentation

Segmentation is the process in which an image is divided into its constituent objects or parts, and background. It is often the most vital *and* most difficult step in an image analysis task. The segmentation result usually determines eventual success of the analysis. For this reason, many segmentation techniques have been developed, and there exist almost as many segmentation algorithms as there are segmentation problems. The construction of a segmentation algorithm can be thought of as defining a model of the objects that we want to detect in the image. This model is then the basis for the segmentation algorithm.

2.1 Thresholding

In the simplest case, we create a model that says that objects are brighter than the image background, and individual objects are well separated from each other. If we can find a suitable intensity threshold that separates the bright objects from the dark background, the segmentation is completed. We simply find all connected pixels brighter than the threshold, and say that they are our objects, as shown

in Fig. 2, left, where all pixels above a given threshold (in this case intensities >50) are assigned the maximum value (white) while other pixels are assigned the minimum value (black). The tricky part is to find a suitable threshold. There are many different automated thresholding methods, see [29, 30] for a review. One approach is to look for valleys in the image histogram. Plotting the number of pixels per intensity-level against intensity-level creates an image histogram. If objects are bright and background is dark, the histogram will have one peak for objects and one for background, and a valley will be present between the peaks. Figure 3a shows an image of fluorescence labeled nuclei of cultured cells. The image histogram is shown in Fig. 3b, and a threshold is placed at intensity 30. In Fig. 3c, the intensity variation along a row in (a) is plotted against x-position, and the threshold is shown as a horizontal line. The result of thresholding the image at this level, and labeling the different connected components, is shown in Fig. 3d. Clustered objects will not be separated by simple intensity thresholding.

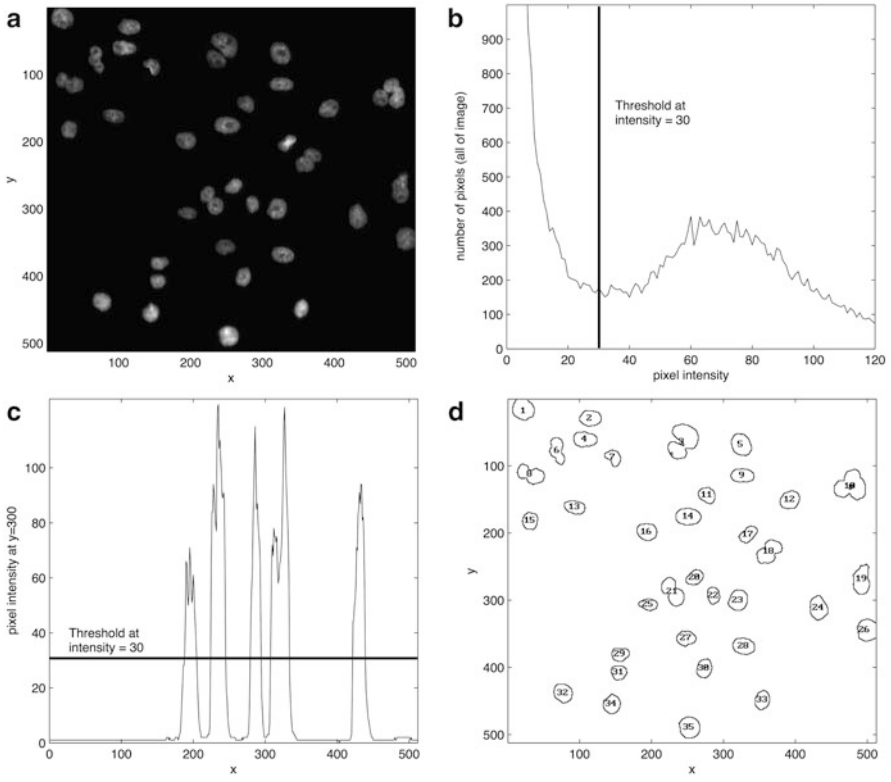

Fig. 3 Image segmentation by thresholding. (**a**) Fluorescence stained nuclei of cultured cells. (**b**) Image histogram of (**a**). A threshold is placed where the histogram shows a local minimum. The *vertical line* corresponds to a threshold at intensity 30. (**c**) An intensity profile along the row y = 300 of (**a**), with the intensity threshold represented by a horizontal line. (**d**) The result after thresholding and labeling of connected pixels. Note that not all nuclei are separated by thresholding

2.2 Watershed Segmentation

If all the objects are brighter than the image background, but clustered, as in the image of cytoplasms in Fig. 4a, thresholding will only separate the objects from the image background, and not separate the individual objects from each other. There is no single threshold that will separate all cells and at the same time find all cells. We can, however, create a model that says that objects have high intensity, and are less intense at borders towards other objects. If image intensity is thought of as height, the cells can be thought of as mountains separated by valleys in an intensity landscape, see Fig. 4b. The segmentation task is then to find the mountains in the landscape.

A segmentation algorithm that has proven to be very useful for many areas of image segmentation where landscape-like image models can be used is watershed segmentation. The method was originally suggested by Digabel and Lantuéjoul, and extended to a more general framework by Beucher et al. [2]. Watershed segmentation has then been refined and used in many situations; see, e.g., Meyer and Beucher [22] or Vincent [34] for an overview. The watershed algorithm works through intensity layer by intensity layer and splits the image into regions similar to the drainage regions of a landscape. If the intensity of the image is thought of as height of a landscape, watershed segmentation can be described as submerging the image landscape in water, and allowing water to rise from each minimum in the landscape. Each minimum will thus give rise to a catchment basin, and when the water rising from two different catchment basins meet, a watershed, or border, is built in the image landscape. All pixels associated with the same catchment basin are assigned the same label. Watershed segmentation can be implemented with sorted pixel lists [35] so that essentially only one pass through the image is required. This implies that the segmentation can be done very fast.

In the case where we want to find bright mountains separated by less bright valleys, we simply turn the landscape up-side-down, inverting the image, and think

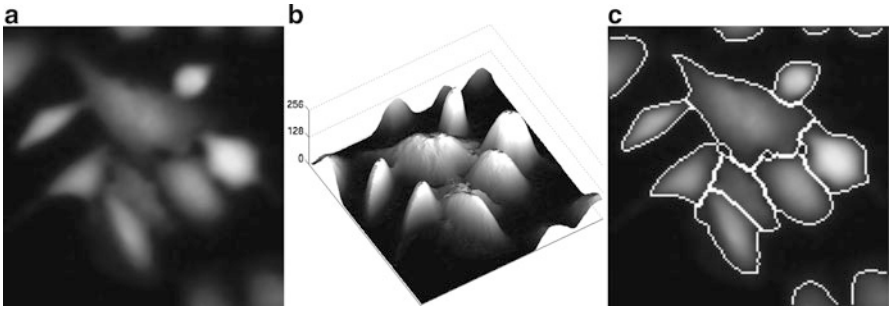

Fig. 4 Image segmentation by watershed segmentation. (**a**) Fluorescence stained cytoplasms of cultured cells. (**b**) The intensity of (**a**) plotted as a landscape. (**c**) The result of watershed segmentation of the inverted image

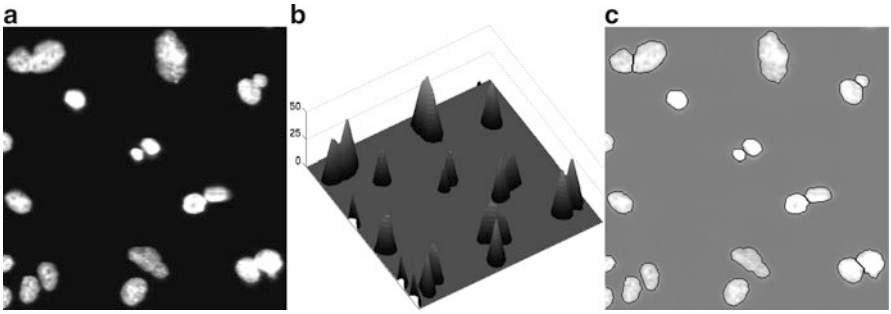

Fig. 5 Shape-based watershed segmentation. (**a**) Free and clustered cell nuclei. (**b**) Distance transformation applied to a thresholded version of (**a**). The distance from each object pixel to the image background is coded as intensity and displayed as height in a landscape. (**c**) The result of watershed segmentation of (**b**) together with (**a**)

of the mountains as lakes separated by ridges instead of mountains separated by valleys. The result after applying watershed segmentation to the image of the cytoplasms can be seen in Fig. 4c.

2.3 Shape-Based Watershed Segmentation

If the clustered objects are not separated by less intense borders, they may have some other feature, or combination of features, that can be included in the segmentation model. One example of such a feature is roundness. The cell nuclei in Fig. 5a are all fairly round in shape, but have internal intensity variations that are sometimes greater than those between the individual nuclei. The clustered nuclei can easily be separated from the background using thresholding. The thresholded image can then be transformed into a distance image, where the intensity of each object pixel corresponds to the distance to the nearest background pixel. Calculation of the distance transformation is very fast, calculated by two passes through the image [3, 4]. The result will be an image showing bright cones, each corresponding to a round object, see Fig. 5b. Watershed segmentation can then be applied to the inverted distance image, and the clustered objects are separated based on roundness, see result in Fig. 5c. Shape-based segmentation has proven useful for segmentation of cell nuclei in a number of studies [13, 20, 24, 27].

2.4 Edge-Based Watershed Segmentation

Intensity variations in the image background often make it difficult to separate the objects from the image background using thresholding. In some cases, it is possible

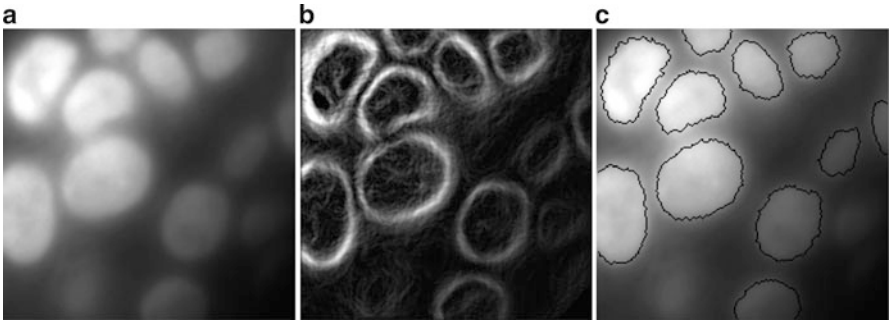

Fig. 6 Edge-based watershed segmentation. (**a**) Fluorescence stained cell nuclei in a section from tumor tissue. Due to background variation, separation of nuclei and background by thresholding is not possible. (**b**) The gradient magnitude of (**a**). (**c**) Result after applying watershed segmentation to the gradient magnitude and overlaying the result with the original image

to reduce these background variations by pre-processing steps that computationally reduce these variations [15, 26]. In other cases, a more advanced model saying that the transition between objects and background is marked by a fast change in image intensity may be applied. In Fig. 6a, the background pixels in the upper left corner of the image have the same intensity as the object pixels in the lower right corner of the image. The objects are still visually clearly detectable as their local intensity is different from the local background.

Intensity changes can be described as the magnitude of the image gradient. The magnitude of the gradient expresses the local contrast in the image, i.e., sharp edges have a large gradient magnitude, while more uniform areas in the image have a gradient magnitude close to zero. The local maximum of the gradient amplitude marks the position of the strongest edge between object and background. The commonly used Sobel operators [32] are a set of linear filters for approximating gradients in the x, y (and z) directions of an image. Adding the absolute values of the convolutions of the image with the different Sobel operators approximates the gradient magnitude image. Figure 6b shows the gradient magnitude, where large magnitude is shown as high image intensity. If watershed segmentation is applied to the gradient magnitude image, the water will rise and meet at the highest points of the ridges, as shown in Fig. 6c. This corresponds to the location of the fastest change in intensity, just as in our segmentation model.

2.5 Merging

When watershed segmentation is applied to an image, water will rise from every minimum in the image, i.e., a unique label will be given to each image minimum. In many cases, not all image minima are relevant. Only the larger intensity variations mark relevant borders of objects. This means that applying watershed segmentation

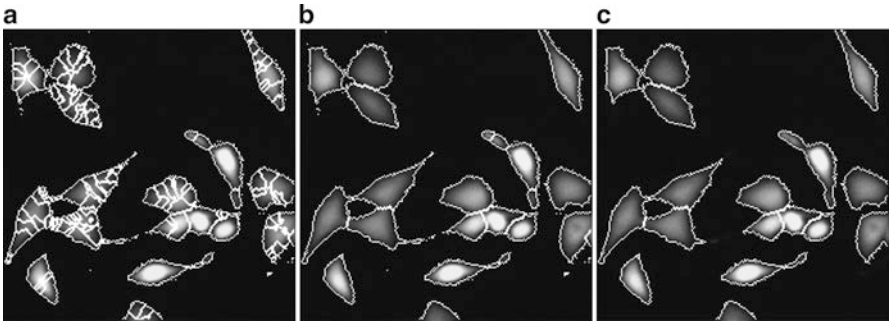

Fig. 7 Edge-based merging. (**a**) Fluorescence labeled cytoplasms with internal intensity variations leading to over-segmentation. (**b**) Result after merging on minimum height of separating ridge. Some over-segmentation still remains. (**c**) Result after further merging of all small objects with the neighbor towards which it has its weakest ridge

will lead to over-segmentation, i.e., objects in the image will be divided into several parts, see Fig. 7a. Over-segmentation can be reduced by a pre-processing step reducing the number of local image minima, e.g., by smoothing the image with a mean or median filter. Smoothing may, however, remove important structures, such as edges, in the image. An alternative to pre-processing is post-processing. After applying watershed segmentation, over-segmented objects can be merged.

Merging can be performed according to different rules, based on the segmentation model. One example is merging based on the height of the ridge separating two catchment basins, as compared to the depth of the catchment basins. The model says that a true separating ridge must have a height greater than a given threshold. All pairs of lakes that at some point along their separating ridge have a height lower than the threshold are merged. The result of merging Fig. 7a at height 10 is shown in Fig. 7b.

Other merging criteria may also be used. For example, if we know that an object must have a certain size, we can include this in our model and say that every object smaller than this size should be merged with one of its neighbors. If there are several neighbors to choose from, we say that merging should be with the neighbor towards which the small object has, e.g., its weakest ridge [37]. The result of this merging method applied to Fig. 7b is shown in (c). The length of the border between two objects can also been used to decide if neighboring objects should be merged or not [33]. Defining the strength of a border as the weakest point along the border may lead to merging of many correctly segmented objects due to single weak border pixels or weak border parts originating from locally less steep gradients. Another simple measure, which is less sensitive to noise and local variations, is the mean value of all pixels along the object border [38].

2.6 Seeded Watershed Segmentation

Including a priori information in our model before applying watershed segmentation can also reduce both over- and under-segmentation. Seeded watershed segmentation [1, 22, 34] means that starting regions, called seeds, are given as input to the watershed segmentation. Water is then only allowed to rise from these seeded regions, and the water rising from the seeds floods all other image minima. The water will continue to rise until the water rising from one seeded region meets the water rising from another seeded region, or a pre-defined object/background threshold. This means that we will always end up with exactly as many regions as we had input seeds.

Seeds can be set manually [19], or in an automated way. For example, we may know that, despite variations in both object and background, each object has a certain contrast compared to its local neighborhood. Such regions can be detected using morphological filters. One example is the extended h-maxima transform, which filters out the relevant maxima using a contrast criterion [31]. All maxima are compared to their local neighborhood, and only those maxima greater than a given threshold h are kept. A low h will result in many seeds, often more than one seed per object. A high h will result in fewer seeds, and some objects may not get a seed at all. An example is shown in Fig. 8. The intensity along a pixel row in Fig. 8a is shown in (b), and the h-maxima are marked in gray. Note that maxima that do not contain gray markers do so in a different image row. Despite background variation, h-maxima are found in all cells, as shown in (c). Seeded watershed segmentation is very useful if we perform our segmentation in the gradient magnitude of the image. We can find seeds in object and background regions based on intensity information in the original image, and then let the water rise from these seeds placed in the gradient magnitude image. Object and background seeds are shown in Fig. 8d and the result after watershed segmentation is shown in (e). The result of this segmentation approach, combined with merging based on edge-strength is shown in Fig. 8f.

Seeds may also come from a parallel image. Cells often vary very much in shape and size, and touch each other. Watershed segmentation will not always give a satisfactory result, as seen in Fig. 9a. If we have a single seed per cell, the task of finding the borders of the cell is greatly simplified. The h-maxima transformation results in useless seeds due to great intensity variations within the cells. The nice thing about cells is, however, that each cytoplasm has a natural marker that may be included in the segmentation model: the nucleus. If the nuclei, which are fairly round in shape and usually nicely separated, are stained and imaged in parallel with the cells, they can be used as seeds for watershed segmentation of the cells. The nuclei of the cells in Fig. 9a are shown in (b), and the result of seeded watershed segmentation using the nuclei as seeds is shown in (c).

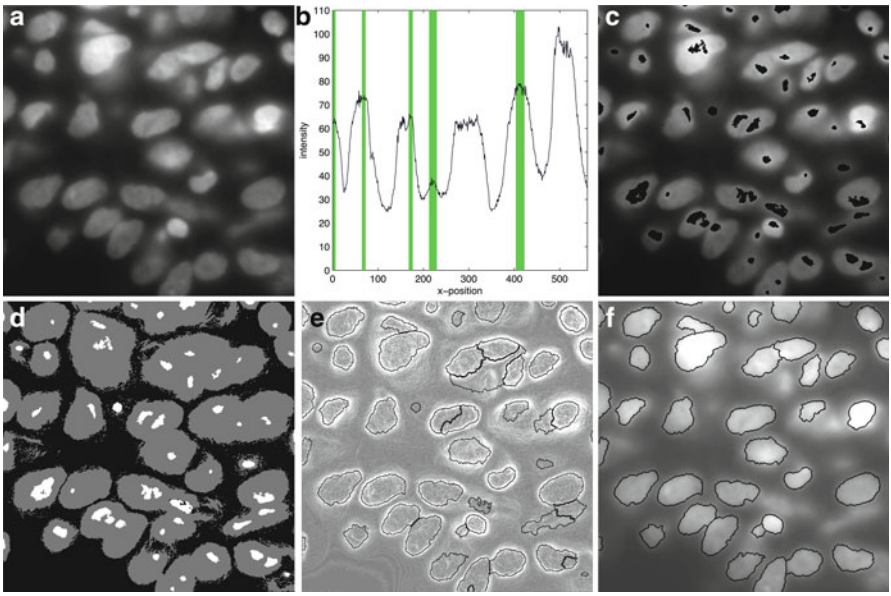

Fig. 8 Seeded watershed segmentation. (**a**) Fluorescence stained cell nuclei in tumor tissue. Due to background variation, separation of nuclei and background by thresholding is not possible. (**b**) Intensity profile across one row of pixels of (**a**), and h-maxima at $h = 5$ shown as *vertical bars* in *gray*. Nuclei without h-maxima have h-maxima in a different row. (**c**) The original image with the h-maxima overlaid. (**d**) Object seeds found by h-maxima transformation (*white*) and background seeds found by h-minima transformation of the gradient magnitude image of (**c**) followed by removal of small objects (*black*). (**e**) Result after seeded watershed segmentation of the gradient magnitude image. More than one seed per object leads to over-segmentation. (**f**) Merging on edge strength reduces over-segmentation. Also poorly focused objects may be removed by this step

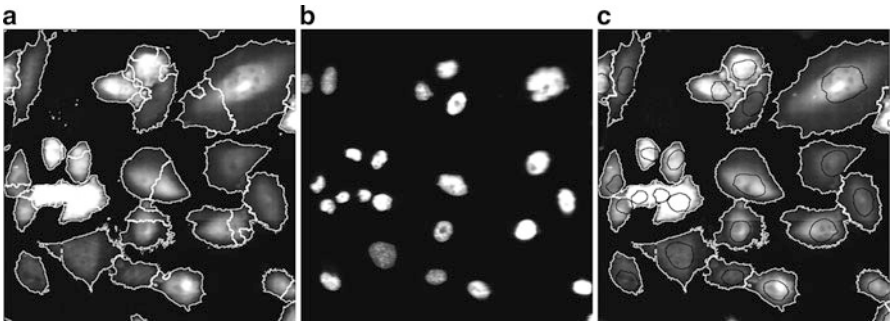

Fig. 9 Cell segmentation using nuclei as seeds. (**a**) Clustered fluorescence labeled cells with varying shapes and intensities are difficult to separate from each other. Watershed segmentation will result in both over- and under-segmentation (*white lines*). (**b**) A parallel image showing the cell nuclei can be used as a seed for watershed segmentation of the cells. (**c**) The result of watershed segmentation (*white lines*) using the nuclei (*black lines*) as seeds

2.7 Extension to Volume Images and Time-Lapse Experiments

Most of the discussed methods can be extended to volume (three dimensional) images. For most methods, the only difference is that instead of working with two-dimensional pixel neighborhoods, we work with three-dimensional voxel neighborhoods. For example, seeded watershed segmentation can be applied to three-dimensional images of fluorescence stained cell nuclei in tumor samples [33, 34]. Here, the 26 side-, edge- and corner- neighbors surrounding each pixel (or voxel) in 3D are considered in each step, starting with finding h-maxima to the final merging of weak edges.

In time-lapse experiments, nuclear stains are often undesirable as they may interfere with the natural behavior of living cells. A nuclear stain may, however, still be used for image segmentation if the cells are fairly stationary. The stain is simply added after the completed experiment, and the same image of the nuclei is thereafter used for segmentation of all time-lapse images [16]. If the cells move, an image of the nuclei can be used as a starting point for backtracking of cell motion.

2.8 Other Segmentation Methods

Many other models for cell segmentation exist, such as iteratively refined active shape models [11] and snake algorithms [25]. A comprehensive review of cell segmentation approaches can be found in Meijering [21]. However, the approaches described here (and summarized in [38]) are used in a wide range of commercially available software for microscopy image analysis, and they are also available in a range of free and open source software, as reviewed by Eliceiri et al. [8]. Free and open source solutions make it easy to share methods between labs, and are valuable for the reproducibility of research. The methods described here are fast and allow large scale screening studies.

3 Feature Extraction and Classification in Large-Scale Image Based Biomedical Screening

Once the objects of interest have been segmented from each other and from the image background, a large number of descriptive features can be extracted from the individual objects in the image, see [28] for an overview. Features may include object size, shape, distribution of sub-structures such as membranes and signals from specific molecular detection methods, local intensity patterns (texture), number of protrusions, number of nearby neighboring cells etc. All features that we can extract are based on the actual pixel values and their spatial arrangements within the object. Features may also include relationships between objects (such as number

of neighbors within a fixed distance), or relationships between larger objects (e.g. cells) and sub-objects (e.g. organelles).

Morphometric, or shape features are features that are based solely on the spatial arrangements of pixels/voxels and include for example area, which is the number of pixels belonging to the object. Perimeter is another morphometric feature, defined as the sum of steps taken when walking around the edge pixels of a 2D object [6, 14], and compactness index is a measure of how compact the object is, described as the object perimeter squared divided by the area. Densitometric, or intensity features are features that describe the gray-level values (without considering the spatial distribution), and include for example the integrated intensity, which is the sum of the intensity values of all pixels/voxels belonging to the object, and the mean object intensity. Another group of features are the more complex textural or structural features that combine spatial and gray-level information. Examples are the gray-level co-occurrence measurement [23] and mass displacement, which is the distance between the center of mass given by the gray-level image of the object and the center of mass given by a binary mask of the object.

The numerical data produced by feature extraction may not always be the desired end result, and it may be difficult to interpret. A simple example is if the goal of the analysis is to decide the percentage of small, medium-sized and large objects in an image. Numerical data representing the area of each object does not provide the final answer. The numerical data has to be analyzed, and each object has to be classified as small, medium, or large in order to calculate the desired percentages. In many cases, the goal of the analysis is to retrieve more complicated information from the images, and a single feature like area is not sufficient for object description and classification. Figure 10 illustrates an example where cells are segmented, features are extracted, and cells are classified into three different classes based on the two most discriminatory feature measurements. If the phenotypes of interest are complex, a larger number of features may be needed, and the different classes have to be separated by multivariate statistic analysis. It is very important to note that increasing the number of features for object classification will not necessarily improve the classification result. It has been observed that, beyond a certain point, inclusion of additional features leads to worse rather than better performance [7]. One should instead try to pick a limited set of features that can discriminate between the relevant populations as well as possible, or use automated feature selection methods. In fact, a very efficient way of automated selection and reduction of features is by iterative feedback and machine learning [10].

Image processing and analysis is often used in high-content analysis/high-throughput screening (HCA/HTS) experiments, searching large libraries of chemical or genetic perturbants, to find new treatments for a disease or to better understand disease pathways. Automated image segmentation, processing and analysis has also more recently shown to be a powerful tool for grouping chemical compounds based on their mechanism of action [17] and to predict the potential performance of novel chemicals in compound libraries [36]. Large-scale experiments analyzed by automated methods require robust models for object detection, such as the ones described here. Robust and diverse staining approaches also increase the chance of

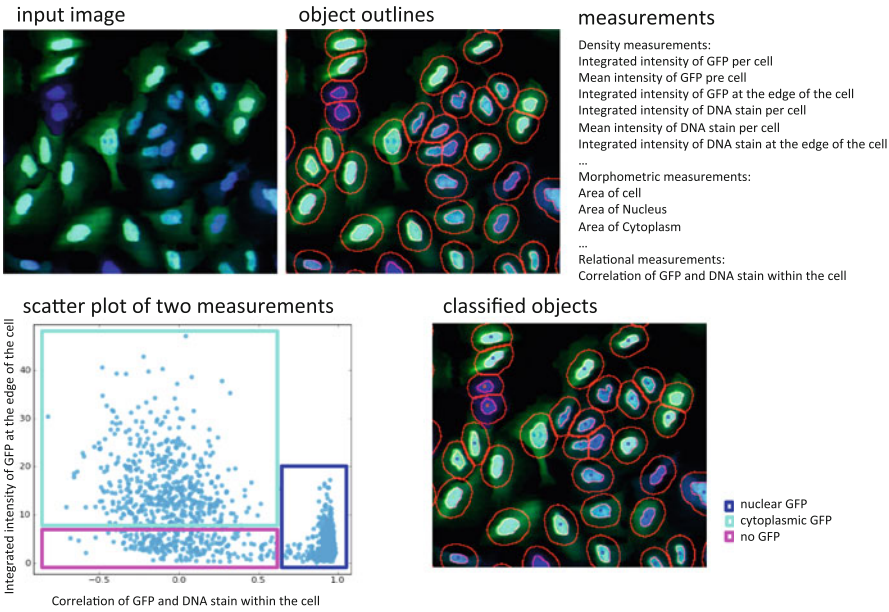

Fig. 10 Illustration of object classification based on feature measurement. The image comes from a translocation assay (image set BBBC013v1 provided by Ilya Ravkin, available from the Broad Bioimage Benchmark Collection, [18]) where one can observe a cytoplasm to nucleus translocation of the Forkhead (FKHR-EGFP) fusion protein in stably transfected human osteosarcoma cells, U2OS. In proliferating cells, FKHR is localized in the cytoplasm. Even without stimulation, Forkhead is constantly moving into the nucleus, but is transported out again by export proteins. Upon inhibition of nuclear export, FKHR accumulates in the nucleus. In this assay, export is inhibited by blocking PI3 kinase/PKB signaling by incubating cells for 1 h with Wortmannin. Nuclei are stained with DRAQ, a DNA stain. The goal is to classify cells into three phenotypes; cells with nuclear GFP (i.e. FKHR-EGFP), cytoplasmic GFP, or no GFP expression. Input images are segmented into cell nuclei and surrounding cytoplasm, where the cytoplasm is defined as any pixels within a fixed distance of a nucleus. Next, a large number of different feature measurements are extracted from each cell. In the classification step, cells are classified into the three classes nuclear GFP (*blue box*), cytoplasmic GFP (*green box*), or no GFP expression (*magenta box*) based on two feature measurements, namely integrated intensity of GFP at the edge of the cell and correlation of GFP and DNA stain within the cell. The distribution of feature measurements is shown as a scatter plot, and the final classification result as small colored squares overlaid the segmented cells

detecting subtle changes of cellular states [9]. Different methods for quantification of image quality are also desirable, especially if the number of images is very large, and system failures such as errors in autofocusing, image saturation and debris may introduce errors in the final screening results [5].

In the most common case, cultured cells model biological processes and disease pathways. Studying disease by culturing cells allows for efficient analysis and exploration. However, many diseases and biological pathways can be better studied in whole animals—particularly diseases that involve organ systems and multicellular

interactions, such as metabolism and infection. The worm *Caenorhabditis elegans* (*C. elegans*) is a well-established and effective model organism, used by thousands of researchers worldwide to study complex biological processes. Samples of *C. elegans* can be robotically prepared and imaged by high-throughput microscopy, just as with cells, mathematical models of worm shape and appearance are required for efficient analysis [39, 40].

Automated analysis of cells as well as model organisms are typical examples where biological questions and their real-life constraints together with the possibilities and limitations of mathematical models require expertise from biologists as well as mathematicians. Image based research as such requires different scientific communities to closely engage and communicate, and there is broad potential for new discoveries in this fast growing field of science.

References

1. Beucher, S.: The watershed transformation applied to image segmentation. Scanning Microsc. **6**, 299–314 (1992)
2. Beucher, S., Lantueéjoul, C.: Use of watersheds in contour detection. In: International Workshop on Image Processing: Real-Time and Motion Detection/Estimation, Rennes (1979)
3. Borgefors, G.: Distance transformations in digital images. Comput. Vis. Graphics Image Process. **34**, 344–371 (1986)
4. Borgefors, G.: On digital distance transforms in three dimensions. Comput. Vis. Image Underst. **44**(3), 368–376 (1996)
5. Bray, M.A., Fraser, A.N., Hasaka, T.P., Carpenter, A.E.: Workflow and metrics for image quality control in large-scale high-content screens. J. Biomol. Screen. **17**(2), 266–274 (2012)
6. Dorst, L., Smeulders, A.W.M.: Length estimators for digitized contours. Comput. Vis. Graph. Image Process. **40**, 311–333 (1987)
7. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
8. Eliceiri, K.W., Berthold, M.R., Goldberg, I.G., Ibáñez, L., Manjunath, B.S., Martone, M.E., Murphy, R.F., Peng, H., Plant, A.L., Roysam, B., Stuurman, N., Swedlow, J.R., Tomancak, P., Carpenter, A.E.: Biological imaging software tools. Nat. Methods **9**(7), 697–710 (2012)
9. Gustafsdottir, S.M., Ljosa, V., Sokolnicki, K.L., Anthony Wilson, J., Walpita, D., Kemp, M.M., Petri Seiler, K., Carrel, H.A., Golub, T.R., Schreiber, S.L., Clemons, P.A., Carpenter, A.E., Shamji, A.F.: Multiplex cytological profiling assay to measure diverse cellular states. PLoS One **8**(12), e80999 (2013)
10. Jones, T.R., Carpenter, A.E., Lamprecht, M.R., Moffat, J., Silver, S.J., Grenier, J.K., Castoreno, A.B., Eggert, U.S., Root, D.E., Golland, P., Sabatini, D.M.: Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc. Natl. Acad. Sci. U. S. A. **106**(6), 1826–1831 (2009)
11. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. **1**, 321–331 (1988)
12. Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., Nilsson, M.: In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods **10**, 857–860 (2013)
13. Krtolica, A., Ortiz de Solorzano, C., Lockett, S., Campisi, J.: Quantification of epithelial cells in coculture with fibroblasts by fluorescence image analysis. Cytometry **49**, 73–82 (2002)
14. Kulpa, Z.: Area and perimeter measurement of blobs in discrete binary pictures. Comput. Graph. Image Process. **6**:434–454 (1977)

15. Likar, B., Maintz, J.B., Viergever, M.A., Pernus, F.: Retrospective shading correction based on entropy minimization. J. Microsc. **197**(3), 285–295 (2000)
16. Lindblad, J., Wählby, C., Bengtsson, E., Zaltsman, A.: Image analysis for automatic segmentation of cells and classification of Rac1 activation. Cytometry A. **57**(1), 22–33 (2004)
17. Ljosa, V., Caie, P.D., Ter Horst, R., Sokolnicki, K.L., Jenkins, E.L., Daya, S., Roberts, M.E., Jones, T.R., Singh, S., Genovesio, A., Clemons, P.A., Carragher, N.O., Carpenter, A.E.: Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J. Biomol. Screen. **18**(10), 1321–1329 (2013)
18. Ljosa, V., Sokolnicki, K.L., Carpenter, A.E.: Annotated high-throughput microscopy image sets for validation. Nat. Methods **9**(7), 637 (2012)
19. Lockett, S.J., Sudar, D., Thompson, C.T., Pinkel, D., Gray, J.W.: Efficient, interactive, and three-dimensional segmentation of cell nuclei in thick tissue sections. Cytometry **31**, 275–286 (1998)
20. Malpica, N., Ortiz de Solorzano, C., Vaquero, J.J., Santos, A., Vallcorba, I., Garcia-Sagredo, J.M., del Pozo, F.: Applying watershed algorithms to the segmentation of clustered nuclei. Cytometry **28**(4), 289–297 (1997)
21. Meijering, E.: Cell segmentation: 50 years down the road. IEEE Signal Process. Mag. **29**, 140–145 (2012)
22. Meyer, F., Beucher, S.: Morphological segmentation. J. Vis. Commun. Image Represent. **1**(1), 21–46 (1990)
23. Nielsen, B., Albregtsen, F., Danielsen, H.E.: Statistical nuclear texture analysis in cancer research: a review of methods and applications. Crit. Rev. Oncog. **14**, 89–164 (2008)
24. Ortiz de Solorzano, C., Garcia Rodriguez, E., Jones, A., Pinkel, D., Gray, J., Sudar, D., Lockett, S.: Segmentation of confocal microscope images of cell nuclei in thick tissue sections. J. Microsc. **193**, 212–226 (1999)
25. Park, J., Keller, J.M.: Snakes on the watershed. IEEE Trans. Pattern Anal. Mach. Intell. **23**(10):1201–1205 (2001)
26. Piccinini, F., Lucarelli, E., Gherardi, A., Bevilacqua, A.: Multi-image based method to correct vignetting effect in light microscopy images. J. Microsc. **248**(1), 6–22 (2012)
27. Ranefall, P., Wester, K., Bengtsson, E.: Automatic quantification of immunohistochemically stained cell nuclei using unsupervised image analysis. Anal. Cell. Pathol. **16**, 29–43 (1998)
28. Rodenacker, K., Bengtsson, E.: A feature set for cytometry on digitized microscopic images. Anal. Cell. Pathol. **25**, 1–36 (2003)
29. Sahoo, P.K., Soltani, S., Wong, A.K.C., Chen, Y.C.: A survey of thresholding techniques. Comput. Vis. Graph. Image Process. **41**, 233–260 (1988)
30. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging **13**(1), 146–165 (2004)
31. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer-Verlag, Berlin Heidelberg (1999)
32. Sonka, M., Hlavac, V., Boyle, R.: Image Processing Analysis and Machine Vision, 2nd edn. Brooks/Cole Publishing Company, Pacific Grove (1999)
33. Umesh Adiga, P.S., Chaudhuri, B.B.: An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images. Pattern Recogn. **34**, 1449–1458 (2001)
34. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans. Image Process. **2**(2), 176–201 (1993)
35. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. **13**(6), 583–597 (1991)
36. Wawer, M.J., Li, K., Gustafsdottir, S.M., Ljosa, V., Bodycombe, N.E., Marton, M.A., Sokolnicki, K.L., Bray, M.A., Kemp, M.M., Winchester, E., Taylor, B., Grant, G.B., Hon, C.S., Duvall, J.R., Wilson, J.A., Bittker, J.A., Dančík, V., Narayan, R., Subramanian, A., Winckler, W., Golub, T.R., Carpenter, A.E., Shamji, A.F., Schreiber, S.L., Clemons, P.A.: Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc. Natl. Acad. Sci. U. S. A. **111**(30), 10911–10916 (2014)

37. Wählby, C., Lindblad, J., Vondrus, M., Bengtsson, E., Björkesten, L.: Algorithms for cytoplasm segmentation of fluorescence labeled cells. Anal. Cell. Pathol. **24**(2–3), 101–111 (2002)
38. Wählby, C: Algorithms for applied digital image cytometry. PhD thesis Uppsala University, Sweden (2003)
37. Wählby, C., Sintorn, I.-M., Erlandsson, F., Borgefors, G., Bengtsson, E.: Combining intensity, edge, and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. J. Microsc. **215**(1), 67–76 (2004)
39. Wählby, C., Kamentsky, L., Liu, Z.H., Riklin-Raviv, T., Conery, A.L., O'Rourke, E.J., Sokolnicki, K.L., Visvikis, O., Ljosa, V., Irazoqui, J.E., Golland, P., Ruvkun, G., Ausubel, F.M., Carpenter, A.E.: An image analysis toolbox for high-throughput *C. elegans* assays. Nat. Methods **9**(7), 714–716 (2012)
40. Wählby, C., Conery, A.L., Bray, M.A., Kamentsky, L., Larkins-Ford, J., Sokolnicki, K.L., Veneskey, M., Michaels, K., Carpenter, A.E., O'Rourke, E.J.: High- and low-throughput scoring of fat mass and body fat distribution in *C. elegans*. Methods **68**(3), 492–499 (2014)