Segmentation for high-throughput image analysis: watershed masked clustering

Yan, K.; Verbeek, F.J.; Margaria, T.; Steffen, B.

Segmentation for High-Throughput Image Analysis: Watershed Masked Clustering

Kuan Yan and Fons J. Verbeek

Section Imaging and Bioinformatics
Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
{kyan,fverbeek}@liacs.nl

Abstract. High-throughput microscopy imaging applications represent an important research field that is focused on testing and comparing lots of different conditions in living systems. It runs over a limited time-frame and per time step images are generated as output; within the time-range a resilient variation in the images of the experiment is characteristic. Studies represent dynamic circumstances expressed in shape variation of the objects under study. For object extraction, i.e. the segmentation of cells, aforementioned conditions have to be taken into account. Segmentation is used to extract objects from images and from objects features are measured. For high-throughput applications generic segmentation algorithms tend to be suboptimal. Therefore, an algorithm is required that can adapt to a range of variations; i.e. self-adaptation of the segmentation parameters without prior knowledge. In order to prevent measurement bias, the algorithm should be able to assess all inconclusive configurations, e.g. cell clusters. The segmentation method must be accurate and robust so that results that can be trustfully used in further analysis and interpretation. For this study a number of algorithms were evaluated and from the results a new algorithm was developed; the watershed masked clustering algorithm. It consists of three steps: (1) a watershed algorithm is used to establish the coarse location of objects, (2) the threshold is optimized by applying a clustering in each watershed region and (3) each mask is reevaluated on consistency and reoptimized so as to result in consistent segmented objects. The evaluation of our algorithm is realized by testing with images containing artificial objects and real-life microscopy images. The result shows that our algorithm is significantly more accurate, more robust and very reproducible.

Keywords: High-Throughput Imaging, Segmentation, Watershed, Fuzzy C-means clustering, Fluorescence Microscopy, Systems Biology.

1 Introduction

Image segmentation is an image analysis method that separates pixels into characteristic groups. For high-throughput image analysis, image segmentation is quintessential in obtaining precise per-object information that need be analyzed. Generic segmentation methods cannot always obtain optimal results. Often we have to tune

T. Margaria and B. Steffen (Eds.): ISoLA 2012, Part II, LNCS 7610, pp. 25–41, 2012.

the generic segmentation method with heuristics to get the result with which further processing will be possible. The methodology described in this paper is used in a biomedical setting; i.e., in a workflow which measurements are extracted from images to support the understanding of indigenous phenomena in the images. Segmentation is but one step in this workflow towards the understanding of the image content by means of pattern recognition. Notably, such understanding can only be retrieved in the context of the application; often in the process of image segmentation, heuristics from the domain at hand are included. Thus, assuring that the understanding is based on correctly measured features makes the segmentation step crucial.

In this paper, we focus on segmentation in the high-throughput (HT) imaging as applied to the study of cell systems. High-throughput applications are often related with high-content applications, here we restrict to high-throughput as for the understanding of the development of our algorithm it suffices. This study is motivated by the consideration that in the field HT-imaging the current segmentation algorithms perform inadequate. In order to explain the development of a new algorithm we briefly introduce the application field. Subsequently, we introduce typical pitfalls of HT-imaging and thereby formulate requirements that are important to the development of our algorithm. Underlying this introduction is the workflow in HT-imaging that starts with image acquisition and through a pipeline of image processing ends with image understanding achieved through pattern recognition.

1.1 High-Throughput Cell Imaging

The application of HT-imaging is of increasing importance in the study of cell systems. In the past decades, there has been a considerable progress in imaging techniques and molecular engineering. Consequently, this progress has been addressed to make the study of cell systems feasible. Starting point for the imaging is the compound microscope that is adapted to study cell systems *in vitro*. That is, as cell cultures in small containers specifically suitable for microscopy imaging. These systems are studied under a range of different conditions including duplicates and controls in the same experiment. This requires a specific setup which is commonly referred to as a high-throughput screening (HTS) [6, 20, 31]. The aim of a screen is to capture and quantify the unique cellular and/or molecular phenotype of a particular cell line under different conditions. The cells are cultured in a 96-well culture plate [20, 31] in which each well represents an experimental condition; some wells are used for duplicates and control groups.

In order to accomplish a quantification of the phenotype by accurate descriptors, a robust image analysis pipeline must be configured. This pipeline receives time-lapse image sequences as input. These time-lapse sequences are captured with a microscope/camera setup and comprehend a complete HTS experiment. The pipeline includes acquisition, preprocessing, segmentation, object labeling, tracking, measurement and classification [4, 5, 28]. From the results of machine learning, conclusions can be formulated that are meaningful and comprehensive in the context of biology. The critical step in the processing pipeline, however, is the extraction of individual objects, i.e. cells, as precise as possible. This requires a robust segmentation algorithm

that produces accurate and reproducible results over the large amounts of images from the HTS. A typical screen requires large volumes of data to be processed; e.g. an experiment for drug-target discovery produces more than ten 96-well plates. Assuming acquisition of three time-lapse image sequences per well, with a length of 150 frames (5 min/frame over 12 hrs), will then result in over 400K images per experiment. Each image contains 50-200 objects (cells), which, normally, cover approximately 10% of the cell population in a well. So, in the application domain of HTS, we have to take into account the conditions under which images of the objects are acquired. There are limitations that need be accounted for in terms of heuristics in the algorithm that is devised for such applications.

Typically, fluorescence microscopy [26, 29] is the microscope modality in HTS imaging and it requires the application of fluorophores to the object under study, a.k.a. fluorescent staining or 'labeling'. Fluorescent staining is based on a class of dyes that have the capacity to emit light under excitation [26]. The major advantage of fluorescence microscopy, over phase contrast or bright field microscopy is that different stains can be simultaneously applied to functional components or protein complexes within the cell, which may reveal underlying phenotypic correlations between cell migration and protein complex localization. Each specific staining can be visualized in a separate channel of the microscope through the use of bandpass filters. The image capture is achieved with a CCD camera mounted on the microscope [4, 5, 28]. The CCD characteristic of a linear response to the amount of light, even at low doses [29], is important in fluorescence microscopy. The amount of fluorophore, through specific binding per object, however, is subject to per-object variation and experimental bias (cf. Fig. 1).

In the 96-well plate, the substrate of each well to which the cells are adhered, is slightly concave toward the center of the well. This complicates the imaging, and contributes in uneven density distribution in the cells. The lens cannot correct for this, it is a trade-off between higher numerical aperture, i.e. resolution [29], and focal depth. As accuracy in the description of the shape is required, a sufficient numerical aperture needs to be chosen for the imaging thus compromising in focal depth. We have to acknowledge these conditions to contribute to the quality of imaging and henceforth might affect the quality of the segmentation result.

1.2 Evaluation of Segmentation Algorithms in HT-Imaging

As indicated, HTS experiments cover dynamic events and therefore images are acquired in time-lapse. This result in large amounts of images, a typical experiment can account for 100K to 500K images per screen; consequently, computational load has to be considered. Moreover, as conditions may differ, the parameters for segmentation cannot be applied over a global set but have to be determined in local environments. In addition, over all images the algorithm needs to cope with situations that may result in erroneous outcome of the features. The crux of the segmentation algorithm is therefore to prevent errors in the measurements that would otherwise introduce misclassification and misinterpretation. The objective is to find all objects and

extract morphological features/descriptors from these objects. A major complication is that the signal, i.e. the fluorescent labeling, is not evenly strong in all objects (cf. Fig. 1b). This is an experimental flaw that needs to be taken into account.

Given the consideration on the segmentation algorithms, a possible candidate is the fuzzy C-means segmentation algorithm, which is derived from the fuzzy C-means clustering algorithm [3]. Similar to Otsu [23], it provides an intensity threshold that is used to separate background and foreground pixels. The disadvantages are similar to those found in the Otsu algorithm. Therefore, we aim at a local approach of a clustering application and we need to consider methods to regionalize the image. Finding such regions can be dealt with through a seeded watershed approach [2, 14]. Via the combination of fuzzy C-means algorithm and watershed algorithm, we derive an innovative form of segmentation, namely watershed masked clustering (WMC).

The WMC algorithm consists of three steps and at each step the segmentation result is further refined. It first finds several coarse regions; each region is considered a rough mask that requires further optimization. Next, a more precise mask is obtained from each coarse region. In the final step, the masks are assessed and, if necessary, corrected using multiple criteria. Following this principle, the WMC algorithm converts a multimodal optimization problem into a simpler collection of several optimization problems while each is guaranteed unimodal.

The WMC is designed to be a robust and dedicated solution to the particular application of the image segmentation in large high-throughput screens to study cell systems. Compared to currently used segmentation approaches, WMC is very sensitive to regional variation of intensity values in images (cf. Fig. 1b); specifically for images of cells with fluorescent labeling.

The remainder of this paper is organized as follows; in section 2, the structure of the WMC algorithm is explained in detail. Subsequently, the performance of the WMC algorithm is illustrated, at the same time the algorithm is compared to a number of other segmentation approaches, i.e. Otsu [23], Bernsen local adaptive thresholding [13], hysteresis edge-based thresholding [9] and level-set methods [30]. Finally, we present our conclusions and discuss our developments and results in broader context.

2 Watershed Masked Clustering Algorithm

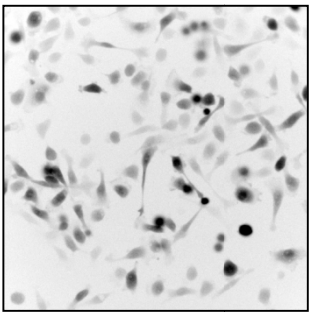
In Figure 2, the outline of the algorithm is depicted. In the discussion of the algorithm, we will refer to this figure and its corresponding details for each of the subparts of the algorithm. For an algorithm to be suitable for high-throughput screening the following requirements must be satisfied:

1. Adaptive to local variations in intensity
2. Capable of processing large amount of images without parameter recalibration
3. Capable of finding a separation between objects
4. Computationally efficient

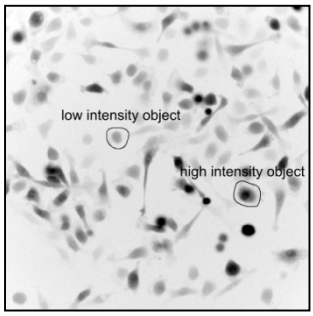
This can be translated to the three main features of the segmentation approach that we propose:

1. Divide the image in intensity regions
2. Find object(s) in each intensity region
3. Check integrity of each object

The intensity regions in the image are established using watershed segmentation. For our algorithm, this is a preprocessing step, as it will not sufficiently separate all objects. Next, in each of the regions a clustering is applied to find the objects with a highest possible precision. In the case, that the preprocessing has split objects that actually should be considered as one object, a correction will follow in the last step of the algorithm. In this step re-evaluation over the objects is applied to get the best possible output. For this re-evaluation, a special form of watershed object segmentation is designed. In Figure 2 the overall pseudo code is given. In the next sections, each of the steps of the algorithm is described in detail.



(a) Fluorescence microscope image of cancer cells with fluorescent label (GFP)



(b) Indication of typical variation in the intensity of the fluorescence in one image

Fig. 1. Image (512x512) from a HTS of cancer cells visualized with a fluorescence microscope (20x). In order to better render intensity differences the inverse LUT of image 1a is used in 1b.

2.1 Region Selection

The implementation of the maxima-seeded watershed masking is based on research by Pinidiyaarachchi [2]. In this flavor of the watershed algorithm, the growing of the watershed region is initialized from a pixel with the highest intensity compared to its neighboring pixels: this particular pixel is referred to as the local maximum. In order to define a valid local maximum, the intensity of such a pixel must exceed the surrounding pixel intensities by a threshold value *h*, where *h* is an estimated level of noise tolerance in terms of intensity (cf. Fig. 3); *h* is commonly referred to as the *h*-maximum [2]. A higher value of *h* provides a less sensitive watershed separation and *vice versa*. In practice, a higher value of *h* often leads to incomplete separation of the objects in the image and moreover, objects that occur in clusters are often not sufficiently separated (cf. Fig. 3b). We can derive the range for the value of *h*, since the *h*-maximum is considered a reference relative to the intensity value of the pixels. Let I_M be the maximum intensity in the dynamic range of the sensor, and I_{max} the maximum intensity in the region under study, the h-maximum is typically in $[1, (I_M-I_{max})]$. In

Figure 3, the results of the maxima-seeded watershed for different values of *h* are depicted. From empirical observations in HTS images (I_M=255), a value *h*=20 provides satisfactory watershed regions.

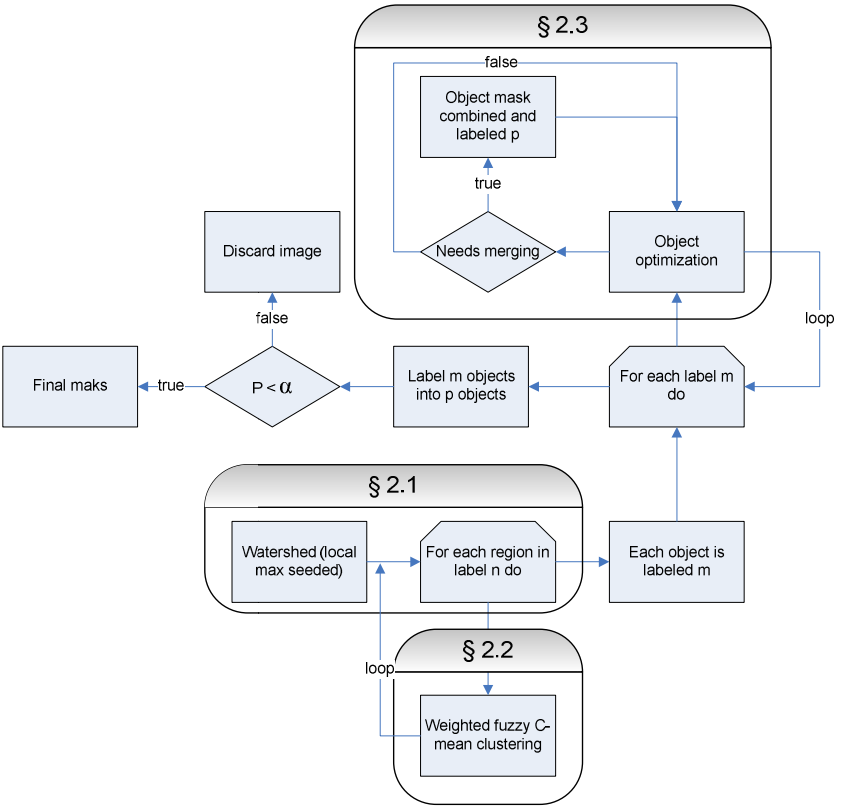
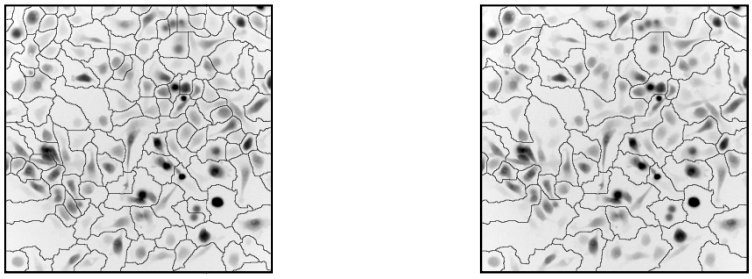


Fig. 2. A workflow diagram illustrates of the three main steps of the Watershed Masked Clustering Algorithm. As part of the automation process, at completion of the loop there is always a quality check Q to prevent wrongly processed images to be part of the analysis.



(a) Original image with watershed lines, h=20 (b) Original image with watershed lines, h=50

Fig. 3. It contains results from the watershed algorithm illustrating the effect of the value of h. The (a) and (b) illustrate the watershed cutting lines of Fig. 1b for h=10.

2.2 Object Segmentation

The application of the watershed method provides coarse regions. Given these coarse regions, starting point for the next step in our algorithm follows from the intrinsic features of the watershed method; it is guaranteed that:

1. In each watershed region, the intensity landscape is always unimodal [2, 14].
2. Seeded watershed implements a restriction on the possible starting point of path searching. An empty region usually does not contain valid seed, thus no watershed region will be formed in an empty region.

In order to find the object in each region, an approach is required that is capable of establishing a local adaptive threshold while being computational finite. Such can be accomplished by a weighted fuzzy C-means clustering algorithm (WFCM). This clustering is applied sequentially in each of the regions to search an optimal value for thresholding in the region. Consequently, each region has its own threshold value taking into account local conditions, i.e. the local variation in image intensity.

In addition, the WFCM method has a set of weighting factors ω that allows the introduction of prior probability of the pixel membership in clusters. The definition of such a weighting factor is similar to the reversed version of the prior probability in Bayesian theory. A smaller weighting factor is assigned to the cluster having, potentially, a larger standard deviation and vice versa. The sum of all weighting factors is always one. The weighting factor ω can be directly derived from the data [15, 21] however, with a known type of image data, commonly, a preset value is used. The WFCM method is formalized as:

$$u_{ij} = \left(\sum_{c}^{k=1} \left(\frac{\omega_j \cdot \left\| x_i - c_j \right\|}{\omega_k \cdot \left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \tag{1}$$

where u_{ij} denotes the membership matrix, c_j is the jth cluster, x_i is the data vector i and ω_j is the weighting factor for cluster j. Empirically, it has been established that for cell imaging a value of ω = 0.2 for the foreground and a value of ω = 0.8 for the background is sufficient. This should be interpreted as: (1) there is an 80% chance a certain pixel is belongs to the foreground and (2) there is a 20% chance that a certain pixel is belongs to the background. By increasing the weighting factor for the foreground, less intense structures, such as cell protrusions or objects with a low overall intensity, will be discarded. In this manner, the weighting factor works similarly to the parameter for the degree of sensitivity in the fuzzy c-means clustering algorithm [21]. Along with eq.1, the clusters are formalized as:

$$c_j = \frac{\sum_{N}^{i=1} \left(u_{ij}^{m} \cdot x_i \right)}{\sum_{N}^{i=1} \left(u_{ij}^{m} \right)}, \tag{2}$$

where u_{ij} denotes is the membership matrix at step k and m is the, so called, fuzzy coefficient that expresses the complexity of the model, by default $m=2$. In our algorithm, we strive at a quick convergence of the WFCM and therefore the initial seeds for c are defined as follows:

$$c_{Foregroundseed} = \bar{I} + (2^{nb} - 1) \cdot \frac{I_{max} - \bar{I}}{\sigma(I)}, \tag{3}$$

$$c_{Backgroundseed} = \bar{I} - (2^{nb} - 1) \cdot \frac{\bar{I} - I_{min}}{\sigma(I)}, \tag{4}$$

where I_{min}, I_{max} denote the minimum/maximum intensity in the image I, $\bar{I}$denotes the mean of the intensities in image I, $\sigma(I)$ denotes the standard deviation in the intensities of the image *I* and *nb* denotes the dynamic range of the intensity expressed in number of bits. In the standard case of unsigned 8-bit images *nb=8*.

This approach provides a robust solution to address the complexity in the HTS images regarding variation in foreground and background intensities. The application of this step results in a binary object in each of the regions of step 1 (cf. §.2.1), if correct, shape features can be derived. However, the watershed method might have introduced some irregularities in the establishment of the coarse regions, which requires an additional evaluation; this evaluation is elaborated in the next section. Examples of the application of this step of the WMC algorithm are worked out in the section 3.

2.3 Object Optimization

At onset of our algorithm, the watershed segmentation is applied resulting in regions that are individually processed. Depending on the variation in the data, the watershed algorithm is known to result in an overcut of the segmentation; overcut is referred to as the situation in which the watershed segmentation produces more regions than actually present in the image [12]. This overcut might affect the individual objects, because of which the objects need be split or merged (cf. Fig. 4). Therefore, the last step in our algorithm is to compensate for the possible overcut caused by the watershedding. This process is an object optimization as we evaluate the results obtained in the object segmentation. In this procedure, only the objects that share a border with a watershed line are evaluated, as these objects are the candidates for overcut.

The solution for the object optimization is a merging mechanism that uses multiple criteria; currently, two criteria are implemented, i.e.:

1. Evaluation of the strength of watershed line; the objects are merged based on a local difference in maximum and average intensity in the object.
2. Evaluation of the orientation of the objects; the object are merges based on assessment of the difference in orientation of their principal axes.

For criterion 1, we implemented an intensity-based merging algorithm so as to estimate the necessity of merging the objects through the evaluation of the strength of the watershed lines. In this function all watershed lines are evaluated. This criterion can be generalized with the evaluation function *K*:

$$K\ (l_i) \rightarrow min\left(\frac{\delta_1}{\tau_1}, \frac{\delta_2}{\tau_2}\right) > T_k, \tag{5}$$

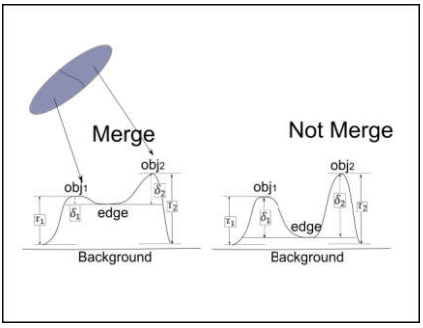
where the l_i denotes the i^{th} watershed line, δ_1 denotes the difference between the average intensity under the watershed line and maximum intensity of object on one side of the watershed and similarly, δ_2 represents the object on the other side of the watershed

line; where τ_1 and τ_2 denote the difference between the maximum and minimum intensity value within one object on either side of the watershed line l_i. A valid watershed line should fulfill the condition given in eq. 5. If $K(l_i)$ exceeds a threshold T_k then the objects on either side of the line are merged to one and the watershed is neglected. In Figure 4a the intensity-based merging criterion is illustrated.

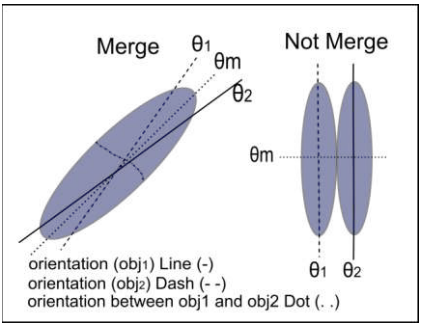
For criterion 2, we implemented an orientation-based merging algorithm [10, 11, 25], which provides a unique possibility to split/merge large structure complexes or elongated objects. At watershed line l_i we consider the principal axis of the objects on either side of the line. A two component Boolean function is designed so that, when true, i.e. both components are true, the objects will be merged and the watershed line will be neglected. This function P is written as:

$$P(l_i) \rightarrow \begin{cases} |\theta_1 - \theta_2| < T_p \\ |\theta_1 - \theta_m| + |\theta_2 - \theta_m| < T_p' \end{cases} \quad (6)$$

where θ_1 denotes the angle between the horizontal image axis (x-axis) and the principle axis of object 1, similarly θ_2 is defined for the object on the other side of watershed line l_i. The θ_m is the angle between the horizontal image axis and the line crossing the centers of mass of the two objects (cf. Fig. 4b). The components in $P(l_i)$ are separately evaluated; so, if the principle axis of each individual object spans a minimum angle T_p while the line crossing the centers of mass of the two objects lies within the angular wedge T_p of the two principle axes, only then these two objects will be merged. In Figure 4b, the orientation-based merging is illustrated by two cases.



(a) Two typical cases of intensity based merging; (left) merge realized and (right) merge not realized using K(li) (eq. 5)



(b) Two typical examples of orientation based merging; (left) merge realized and (right) merge not realized using P(li) (eq. 6)

(c) Sample object

(d) Overcutting from step 1

(e) Merging from step 3

Fig. 4. Illustration of the merging of objects based on a combination of criteria; in (c,d,e) a specific case for one object (cell) is illustrated

Once the object optimization is applied, one can be certain that all objects are correctly extracted and these can be subject to a shape characterization. For the specific case of HTS time-lapse images, both shape and intensity profile of the object are

measured. The intensity profile of an object is derived by applying the binary mask to the original image. A large range of features can be used [25] so that features can be used to discriminate between experimental conditions that are applied [18, 19, 31].

3 Performance of the WMC Algorithm

In this section the performance of our segmentation algorithm will be addressed. In order to get a good impression of its robustness, we employed two tests: i.e., a test with artificially generated images and a test with images from real HTS experiments. Each artificial image contains a number of generated ellipsoid objects. Each HTS image contains in vitro cells that are fluorescently labeled. The performance estimation for each algorithm is derived from the comparison between the binary mask obtained by the algorithm and the corresponding ground-truth binary mask for each image. In the generation of the test images, the ground-truth masks are for the artificial test images are explicitly constructed. The usage of such artificial image provides an image test set with an unbiased ground-truth and controllable noise, allowing the emulation of a worst scenario in fluorescence microscopy imaging.

The pixel-level mismatch of the comparisons is calculated for all algorithms. The rationale behind this test is to simulate the typical data processing workflow for HTS, therefore the parameters used for each of the algorithms are optimized only once and henceforth applied to the whole image set in the experiment. For none of the algorithms in the experiment an individual tuning is applied. The parameters for all algorithms were obtained from the HT screening literature [1, 4, 5, 16, 28].

Segmentation algorithms are often considered simplified versions of linear classifiers trained in intensity space [13, 23, 24]. Similar to the error estimation for a classifier, the error test normally covers both type I (False Positive) and type II (False Negative) errors. Often the performance of a segmentation algorithm is assessed using the number of correct and incorrect segmented pixels [21]. Only covering the type I error may lead to an overtraining of the algorithm [3]. For a balanced conclusion we take into account both type I and type II error. Furthermore, instead of just using the two errors types, we introduce the F1-score [7]. The two types of errors for different algorithms are defined in terms of the true positive and true negative. True positive (TP) is defined as the ratio of pixel overlap between the ground-truth mask and the segmented mask by each algorithm, expressed as:

$$TP = \frac{M \cap M'}{M'}, \tag{7}$$

where M' is the set of pixels belonging to the foreground of binary mask provided by the algorithm and M is the set of pixels belonging to the foreground of the ground-truth mask. In similar fashion, the true negative (TN) is calculated from:

$$TN = \frac{\bar{M} \cap \bar{M}'}{\bar{M}'}. \tag{8}$$

In this way, TP represents the percentage of correctly segmented foreground pixels whereas TN represents the percentage of correctly segmented background pixels. Form the values of TP and TN, the false positives (FP) are derived, i.e. FP = 1-TP (percentage of incorrectly segmented foreground pixels), and likewise the false

negatives (FN) are derived, i.e. FN = 1-TN (incorrectly segmented background pixels). From these values, the sensitivity and specificity the [8] are calculated by:

$$sensitivity = \left(\frac{TP}{TP + FN}\right) \tag{9}$$

$$specificity = \left(\frac{TN}{FP + TN}\right) \tag{10}$$

Given the results, the specificity and the sensitivity for all of algorithms of a particular set of test images can be computed. The results are shown in Table 3. In addition, from the specificity and sensitivity, the F1-score is derived by:

$$F1 = 2 \cdot \frac{specificity \cdot sensitivity}{specificity + sensitivity}. \tag{11}$$

An ideal segmentation algorithm should yield the highest F1-score but this only occurs when both specificity and sensitivity are approaching 100%.

In the next sections, we tested for 5 algorithms, i.e. Bernsen local adaptive thresholding, Otsu thresholding, Level-set segmentation, Hysteresis edge-based segmentation, and our WMC algorithm. We have included Fuzzy C-means clustering (FCM) to the tests to illustrate the enhanced performance of our approach. All algorithms have claimed the intrinsic capacity of performing well under noisy conditions typical to HTS imaging [4, 5, 16, 25, 28]. For the algorithms, open-source plug-ins implementations available in ImageJ [32] and CellProfiler [1] have been used without modifications.

3.1 Artificial Objects and Test Images

The intended application for our segmentation algorithm is high-throughput cell imaging. In order to understand and verify the behavior and performance of our algorithm, ground-truth images with objects resembling the shapes which are normally found in time-lapse cell imaging, are constructed (cf. Fig. 5a). Each image consists of a number of ellipsoid objects and each object has a unique intensity profile. The intensity profile (landscape) is generated through an exponential decay function that is initiated at the centre of each object. The minimum and maximum value of an intensity profile of an object is generated using a uniform distributed random generator and scaled in the range of [20, 255] – in this way sampling to an 8-bit image is simulated. In addition, the orientation of each of the objects is varied by applying a rotation to each of the object in the range of [-30º, 30º] using the center of mass as the pivot; the rotation angle is selected from a uniform random generator. The original binary image with all the objects is kept as the absolute ground-truth mask for the segmentation so that error estimation can be applied over a range of test images that are subjected to a range of different conditions of noise. In this test, a total amount of 30 images is generated. To simulate image noise typical to HTS and fluorescence microscopy, Poisson noise is generated and applied to the images.

3.2 Performance Test with Artificial Images

All algorithms are applied over the same 30 test images (cf. Fig. 5). The F1-scores are listed in Table 1. The object merging accuracy in WMC is also tested using the same

image set. An overcut object is defined as a group of objects obtained by segmentation algorithm share the same object in ground truth mask. A total amount of 238 overcut objects are detected in this image set. Using object optimization, the WMC recovers 202 out of 238 overcut objects, i.e. approximately 85%.

3.3 Microscope Images

In order to test the performance of the WMC in the images it is designed for, we have selected two sets of images from the application domain. The intention of the test is to illustrate the performance in a typical setting and compare the performance with respect to the same selection of segmentation algorithms used in the artificial test images; the test is completely similar to the test with the artificial test images. We will make use of two different image sets. The first image set is known as the "Human HT29 Colon Cancer" dataset [16] (cf. Fig. 6) containing 12 images of human HT29 colon cancer cells. The samples were stained for the nucleus (Hoechst) and the cytoplasm (phalloidin) in two separate channels. The second image set is a time-lapse image sequence, i.e. a dynamic process, which is an MLTn3 cell line [17, 19] used to study migration in live cancer metastasis processes (cf. Fig. 7). It consists of 96 time-lapse image sequences, each of 75 frames. Each sequence portrays an *in vitro* cell migration pattern typical in HTS experiments. The cytoplasm is stained through green fluorescent protein (GFP). For the performance tests, we will only use the first 14 images of the sequence to reduce the size of the image set to reasonable for proportions for this test. In addition, for this image set also a ground-truth image is required. The MTLn3 ground-truth images were obtained by manual segmentation performed by biologists through tracing on a digitizer tablet (WACOM Cintiq). In contrast to the artificial image set, manual segmentation may contain observation bias between and within observers. To that end the manual segmentation is replicated a few times to reduce observer effects.

We will further refer to the first image set as the HT29 set and the second set as the MTLn3 set. The HT29 set is captured at significant higher resolution compared to the MTLn3 set. These two sets are considered a reasonable representation of the scope of the images which are typical input for the WMC algorithm.

3.4 Performance Test with Microscope Images

For the two sets the results are presented in two tables and examples of the segmentation are given in two figures. In Table 2 and 3, the results of the experiment for this set are presented. The sensitivity and specificity are used as the performance indicators. From the result we can conclude that WMC has the best overall performance. It produces stable and robust results for the HT29 set (cf. Table 2). Compared to WMC, the standard FCM algorithm is similar in sensitivity but lower in specificity; the Otsu segmentation is higher in specificity but significantly lower in sensitivity; the Hysteresis segmentation has a similar performance as the WMC. At this point, it is important to realize that the quality of the HT29 set (cf. Fig. 6) is not, in all case, representative for the real high-throughput screens, especially if we consider the dynamic behavior common to live-cell imaging (cf. Fig. 7).

Table 1. Specificity and sensitivity of segmentation efficiency using artificial images

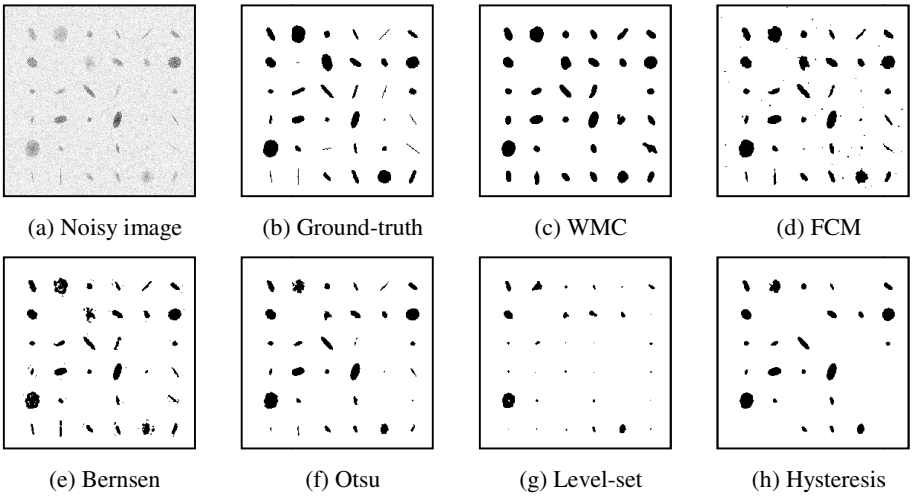
Performance	WMC	FCM	Bernsen	Otsu	Level-set	Hysteresis
Specificity	95.40%	98.11%	78.42%	86.02%	98.38%	79.78%
Sensitivity	96.62%	91.73%	90.14%	98.22%	71.52%	85.16%
F1 Score	95.99%	94.50%	84.60%	91.83%	74.98%	82.95%

Table 2. Specificity and sensitivity of the segmentation algorithms in the HT29 image set

Performance	WMC	FCM	Bernsen	Otsu	Level-set	Hysteresis
Specificity	98.59%	97.81%	99.84%	98.75%	70.68%	97.57%
Sensitivity	98.78%	98.64%	83.10%	89.67%	53.43%	97.88%
F1 Score	98.68%	98.23%	92.11%	93.40%	31.29%	97.73%

Table 3. Specificity and sensitivity of segmentation algorithms in MTLn3 image set

Performance	WMC	FCM	Bernsen	Otsu	Level-set	Hysteresis
Specificity	84.80%	95.76%	99.63%	90.60%	59.10%	82.82%
Sensitivity	91.45%	74.16%	59.25%	80.64%	56.35%	88.13%
F1 Score	88.05%	78.27%	47.63%	83.79%	53.33%	85.81%



(a) Noisy image (b) Ground-truth (c) WMC (d) FCM

(e) Bernsen (f) Otsu (g) Level-set (h) Hysteresis

Fig. 5. (a) noise-added artificial test image, (b) ground-truth masks for the object, (c) to (h) are binary images obtained by corresponding segmentation algorithms. These results are used to compute the errors listed in Table 1.

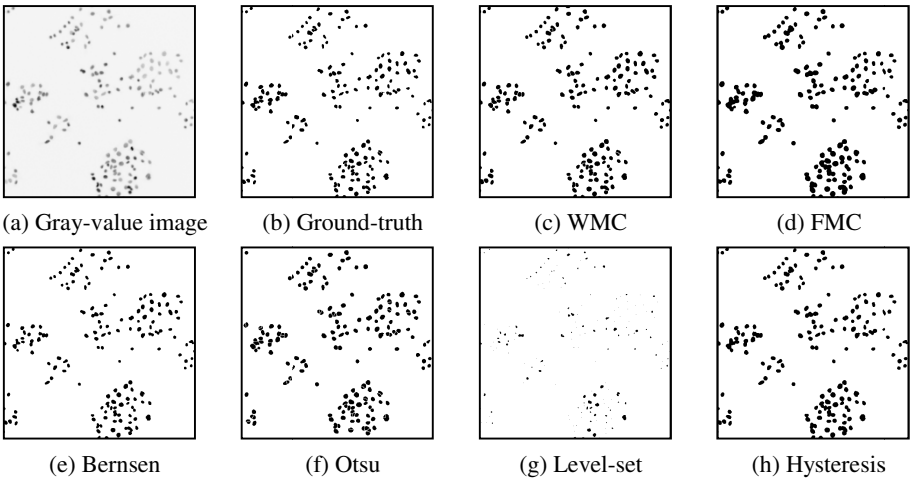


Fig. 6. (a) Original HT29 image acquired with a 10x lens; image size is 512x512 pixels, 8 bit. (b) Ground-truth masks and (c-h) masks obtained by the segmentation algorithms. These results are used to compute the errors listed in Table 2.

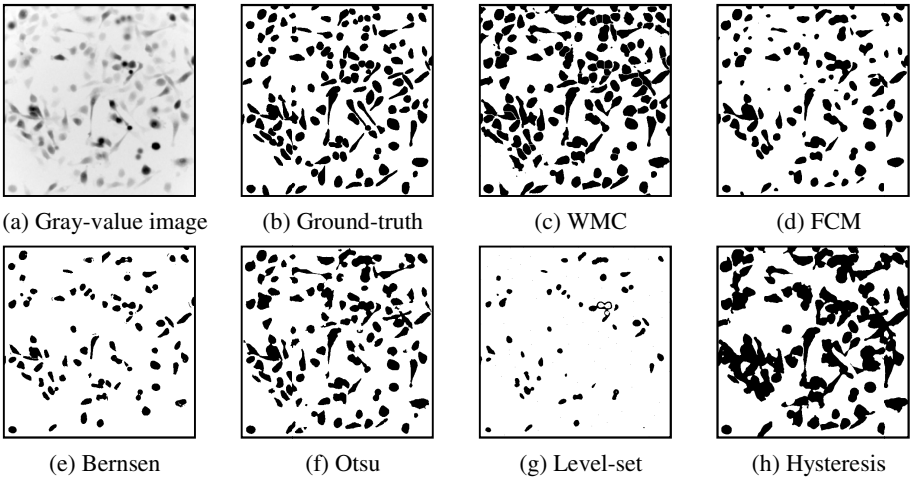


Fig. 7. (a) Original MTLn3 image acquired with a 20x lens (NA 1.4), image size 512x512 pixels, 8-bit (b) ground-truth masks and (c-h) masks obtained by the segmentation algorithms. These results are used to compute the errors listed in Table 3.

The MTLn3 test set is, *de facto*, undersampled in terms of temporal resolution, but a good representation of a HTS with *in vitro* cell migration (cf. Fig. 7a). When finding ground-truth segmentation through manual methods, the variance of the intensities over the whole image set is indicative for the complications that will be faced in automated methods. In Table 3, the results of the experiment with the MTLn3 set are shown. It is immediately clear that the overall performance is much lower compared

to the other test set. However, also in this test the WMC still shows the highest performance. Compared to the WMC algorithm, the FCM algorithm is higher in specificity but significantly lower in the sensitivity. Hysteresis still portrays a good and stable performance. Compared to the previous experiment (cf. Table 2), the WMC algorithm performs quite stable under the different circumstances. In conclusion, from Table 2 and 3 it can be established that the WMC algorithm outperforms the other algorithms, which confirms its applicability for this area of bio-imaging research.

4 Conclusions and Discussion

We have proposed a segmentation algorithm for high throughput imaging that performs better than algorithms that have been used for this purpose so far. The algorithm consists of three steps, a watershed region selection followed by a fuzzy C-means clustering and if necessary followed by a correction for oversegmentation. The algorithm is particularly suitable for imaging in the domain of functional cytomics and high-throughput screenings. We have compared the WMC algorithm with five others and the results of the evaluation convincingly demonstrate the performance WMC algorithm. Over all tests, the WMC algorithm has the best recall (F1-score) without excessive increase in computation time. In the domain of cytomics, analysis is performed *post hoc*; and thus computation time is not a critical component of the analysis but segmentation robustness is. In practice, WMC is now used and we are obtaining high precision results that are understood in biological context [19, 20, 31].

The major advantage of the WMC algorithm is that it can deal with variations in staining intensity typical for bio-imaging and specific to high-throughput *in vitro* experiments. The local intensity variations in the image limit application of Otsu segmentation; it requires a global optimum for the threshold, which may not be possible. Along the same line, the level-set method is not suitable as it presumes a consistent intensity for the objects in the image. The regional approach in WMC followed by a local clustering transforms the segmentation to a local problem so that threshold levels can be found efficiently. For segmentation in cytomics edge based methods are noise susceptible, therefore intensity variations necessitate region based approaches. This is confirmed from our findings comparing Hysteresis segmentation to WMC, especially with more artificial noise (cf. Table 1) or staining variations in the image (cf. Table 3).

The WMC consists of three independent steps and if we consider these individually further improvements can be formulated. In step 1, the watershed algorithm, the initialization of the watershed algorithm is currently based on local maxima; other schemas must be investigated to render a better initialization. Now, *a priori* knowledge is not used whereas this might facilitate a better estimate for the initialization. In step 2, fuzzy weighted C-means clustering is used, however, other clustering approaches can be probed; similarly to step 1, *a priori* knowledge on the intensity distribution might be supportive in finding a better clustering approach. Regarding step 3, we implemented only a few of situations of oversegmentation (cf. §.2.3). This particular step of the algorithm can be adapted to experimental conditions, i.e. *a priori* knowledge can be

tuned with respect to the experiment so as to overcome certain imperfections of earlier steps. In future research this will be elaborated, however, the global idea of the WMC algorithm will stand its case (cf. Table 2 & 3).

The WMC has been successfully applied to other experiments in the domain of bio-imaging, e.g. detection of small vessels [18] and chromosomes. With further generalization, the algorithm can be engaged in a broader scale of imagery. The future research on the tuning of the subsequent steps of the WMC algorithm will contribute to this generalization.

Acknowledgements. This work has been partially supported by the Netherlands' Bioinformatics Centre (NBIC), BioRange Project. The authors like to express their gratitude to our collaborators Dr. S. Le Dévédec, Dr. S. Zovko and Prof. B. van de Water, for making available the MTln3 set. The MTLn3 ground-truth masks are obtained in collaboration with Toxicology/LACDR, Leiden University, the Netherlands. The HT29 was publically made available by the Broad Institute/MIT, USA.

References

1. Carpenter, A., Jones, T., Lamprecht, M., Clarke, C., Kang, I., Friman, O., Al, E.: CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. Genome Biology 7(10) (2006)
2. Pinidiyaarachchi, A., Wählby, C.: Seeded Watersheds for Combined Segmentation and Tracking of Cells. In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, vol. 3617, pp. 336–343. Springer, Heidelberg (2005)
3. Webb, A.: Statistical Pattern Recognition, 2nd edn. Wiley, UK (2005)
4. Neumann, B., Held, M., Liebel, U., Erfle, H., Rogers, P., Pepperkok, R., et al.: High-throughput RNAi screening by time-lapse imaging of live human cells. Nature Methods 3(5), 385–390 (2006)
5. Neumann, B., Walter, T., Hériché, J.K., Bulkescher, J., Erfle, H., Conrad, C., et al.: Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464(7289), 721–727 (2010)
6. Huang, C., Rajfur, Z., Borchers, C., Schaller, M., Jacobson, K.: JNK Phosphorylates paxillin and Regulates Cell Migration. Nature 424, 219–223 (2003)
7. van Rijsbergen, C.: Information Retrieval. Butterworth-Heinemann, UK (1979)
8. Altman, D.G., Bland, J.M.: Statistics Notes: Diagnostic Tests 1: Sensitivity and Specificity. BMJ 308(1552) (1994)
9. Hancock, E., Kittler, J.: Adaptive Estimation of Hysteresis Thresholds. In: Proc. of Computer Vision and Pattern Recognition, pp. 196–201 (1991)
10. Verbeek, F.J.: Three Dimensional Reconstruction from Serial Sections Including Deformation correction. PhD Thesis, Delft University of Technology, The Netherlands (1995)
11. Verbeek, F.J.: Theory & Practice of 3D-reconstructions From serial Sections. In: Baldock, R.A., Graham, J. (eds.) Image Processing, A Practical Approach, pp. 153–195. Oxford University Press, Oxford (1999)
12. Angulo, J., Schaack, B.: Morphological-Based Adaptive Segmentation and Quantification of Cell Assays in High Content Screening. In: Proc. of the 5th IEEE International Symposium on Biomedial Imaging, pp. 360–363 (2008)

13. Bernsen, J.: Dynamic Thresholding of Grey-Level Images. In: Proc. of the 8th Int. Conf. on Pattern Recognition (1986)
14. Roerdink, J.B., Meijster, A.: The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. Fundamenta Informatica, 187–228 (2000)
15. Fan, J., Han, M., Wang, J.: Single Point Iterative Weighted Fuzzy C-means Clustering Algorithm for Remote Sensing Image Segmentation. Pattern Recognition 42(11), 2527–2540 (2009)
16. Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G.: A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124(6), 1283–1298 (2006)
17. Pu, J., McCaig, C.D., Cao, L., Zhao, Z., Segall, J.E., Zhao, M.: EGF receptor Signaling is Essential for Electric-field-directed Migration of Breast Cancer Cells. Journal of Cell Science 120(19), 3395–3403 (2007)
18. Yan, K., Bertens, L., Verbeek, F.J.: Image Registration and Realignment using Evolutionary Algorithms with High resolution 3D model from Human Liver. In: Proc. CGIM 2010 (2010)
19. Yan, K., Le Dévédec, S., van de Water, B., Verbeek, F.J.: Cell Tracking and Data Analysis of in vitro Tumour Cells from Time-Lapse Image Sequences. In: Proc. VISAPP 2009, pp. 281–287 (2009)
20. Damiano, L., Le Dévédec, S.E., Di Stefano, P., Repetto, D., Lalai, R., Truong, H., Xiong, J.L., Danen, E.H., Yan, K., Verbeek, F.J., Attanasio, F., Buccione, R., van de Water, B., Defilippi, P.: p140Cap Suppresses the Invasive Properties of Highly Metastatic MTLn3-EGFR Cells via Paired Cortactin Phosphorylation. Oncogene 30(2) (2011) (in Press)
21. Ma, L., Staunton, R.: A modied fuzzy C-means image segmentation algorithm for use with uneven illumination patterns. Pattern Recognition 40(11), 3005–3011 (2007)
22. Sezgin, M., Sankur, B.: Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. Journal of Electronic Imaging 13(1), 146–165 (2004)
23. Otsu, N.: A Threshold Selection Method from Gray-level Histogram. IEEE Transactions on Systems, Man and Cybernetics 9, 62–66 (1979)
24. Venkateswarlua, N., Raju, P.: Fast Isodata Clustering Algorithms. Pattern Recognition 25(3), 335–342 (1992)
25. van der Putten, P., Bertens, L., Liu, J., Hagen, F., Boekhout, T., Verbeek, F.J.: Classification of Yeast Cells from Image Features to Evaluate. Pathogen Conditions. In: SPIE 6506, MultiMedia Content Access: Algorithms & Systems, vol. 6506, pp. 65060I-1–65060I-14 (2007)
26. Goldman, R., Swedlow, J., Spector, D.: Live Cell Imaging: A Laboratory Manual. Cold Spring Harbor Laboratory Press, USA (2005)
27. Medina-Carnicer, R., Madrid-Cuevas, F., Carmona-Poyato, A., Muñoz Salinas, R.: On candidates selection for hysteresis thresholds in edge detection. Pattern Recognition 42(7), 1284–1296 (2008)
28. Pepperkok, R., Ellenberg, J.: High-throughput Fluorescence Microscopy for Systems Biology. Nature Reviews Molecular Cell Biology 7(9), 690–696 (2006)
29. Inoue, S.: Video Microscopy: the Fundamentals, 2nd edn. Springer, USA (1997)
30. Osher, S.J., Fedkiw, R.P.: Level Set Methods and Dynamic Implicit Surfaces. Springer, USA (2002)
31. LeDévédec, S., Yan, K., de Bont, H., Ghotra, V., Truong, H., Danen, E., Verbeek, F.J., van de Water, B.: A Systems Microscopy Approach to Understand Cancer Cell Migration and Metastasis. Cellular and Molecular in Life Science 67(19), 3219–3240 (2011)
32. Collins, T.: Image J for microscopy. Bio Techniques 43, 25–30 (2007)