Hierarchical Segmentation Based Upon Multi-resolution Approximations and the Watershed Transform

Bruno Figliuzzi(✉), Kaiwen Chang, and Matthieu Faessel

Centre for Mathematical Morphology, Mines ParisTech,
PSL Research University, Paris, France
bruno.figliuzzi@mines-paristech.fr

Abstract. Image segmentation is a classical problem in image processing, which aims at defining an image partition where each identified region corresponds to some object present in the scene. The watershed algorithm is a powerful tool from mathematical morphology to perform this specific task. When applied directly to the gradient of the image to be segmented, it usually yields an over-segmented image. To address this issue, one often uses markers that roughly correspond to the locations of the objects to be segmented. The main challenge associated to marker-controlled segmentation becomes thus the determination of the markers locations. In this article, we present a novel method to select markers for the watershed algorithm based upon multi-resolution approximations. The main principle of the method is to rely on the discrete decimated wavelet transform to obtain successive approximations of the image to be segmented. The minima of the gradient image of each coarse approximation are then propagated back to the original image space and selected as markers for the watershed transform, thus defining a hierarchical structure for the detected contours. The performance of the proposed approach is evaluated by comparing its results to manually segmented images from the Berkeley segmentation database.

Keywords: Watershed transform · Multi-resolution approximations · Hierarchical segmentation · Wavelet transform

1 Introduction

Image segmentation is a classical problem in image processing. Its aim is to provide an image partition where each identified region corresponds to an object present in the image. The watershed transform [1–3] is a popular algorithm based upon mathematical morphology which efficiently performs segmentation tasks. It can easily be understood by making an analogy between an image and a topography relief. In this analogy, the gray value at a given pixel is interpreted as an elevation at some location. The topography relief is then flooded by water coming from the minima of the relief. When water coming from different minima meet at some location, the location is labelled as an edge of the image.

J. Angulo et al. (Eds.): ISMM 2017, LNCS 10225, pp. 185–195, 2017.
DOI: 10.1007/978-3-319-57240-6_15

The watershed algorithm is usually applied to the gradient of the image to be segmented. Each minimum of the gradient therefore gives birth to one region in the resulting segmentation. Due to several factors including noise, quantization error or inherent textures present in the images, gradient operators usually yield a large number of minima. A well-known issue of the watershed algorithm is thus that it usually yields a severely over-segmented image as a result.

To overcome the issue of over-segmentation, a first approach is to apply the gradient operator to images that have previously been filtered. Meyer designed morphological filters, referred to as levelling filters, for this particular task [4,5]. Wavelet based filters have also been used to perform the filtering step. In 2005, Jung and Scharcanski proposed to rely on a redundant wavelet transform to perform image filtering before applying the watershed [6]. The advantage of using the wavelet transform is that its application tends to enhance edges in multiple resolutions, therefore yielding an enhanced version of the gradient. Jung subsequently exploited the multi-scale aspects of the wavelet transform to guide the watershed algorithm toward the detection of edges corresponding to objects of specified sizes [7].

An alternative to overcome the over-segmentation issue is to rely on markers to perform the watershed segmentation. This strategy builds upon the assumption that it is possible to roughly determine the location of the objects of interest to be segmented. The idea is then to perform the flooding from these markers rather than from the minima of the gradient. Another approach that was considered is to select the minima of the gradient according to their importance, by considering for instance h-minima [8,9]. Other approaches including the stochastic watershed rely on stochastic markers that are used to evaluate the frequency at which a contour appear in the segmentation [10]. Finally, following a classical trend in image segmentation [11,12], morphological algorithms have been proposed to perform a bottom-up region merging according to some morphological criteria [13,14].

In this article, our aim is to present a novel method to select markers for the watershed algorithm based upon multi-resolution approximations. The main principle of the method is to rely on the orthogonal wavelet transform to obtain successive approximations of the image to be segmented. The minima of the gradient image of each coarse approximation are then propagated back to the original image space and selected as markers for the watershed transform, thus defining a hierarchical structure for the detected contours. The article is organized as follows. We describe the proposed algorithm and state the main properties of the obtained contours hierarchy in Sect. 2. In Sect. 3, we evaluate the performances of the algorithm on the Berkeley segmentation database. Conclusions are finally drawn in the last section.

2 Multiscale Watershed Segmentation

2.1 Multi-resolution Approximation

A function f from $\mathbb{R}^2$ to $\mathbb{R}$ is said to be square-integrable if and only if the integral

$$\int_{\mathbb{R}^2} |f(x_1, x_2)|^2 \mathrm{d}x_1 \mathrm{d}x_2 \tag{1}$$

is finite. We denote by $\mathbf{L}^2(\mathbb{R}^2)$ the set of square-integrable functions. When equipped with the scalar product

$$< f, g >= \int_{\mathbb{R}^2} f(x_1, x_2) g(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2, \tag{2}$$

it is well-known that $\mathbf{L}^2(\mathbb{R}^2)$ is an Hilbert space of infinite dimension.

Let us first introduce the mathematical notion of multi-resolution approximation [15] which plays a central role in the proposed approach. A multi-resolution of $\mathbf{L}^2(\mathbb{R})$ is a sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $\mathbf{L}^2(\mathbb{R})$ satisfying the following properties

1. $\forall (j,k) \in \mathbb{Z}^2,\ f(.) \in V_j \Leftrightarrow f(. - 2^j k) \in V_j$.
2. $\forall j \in \mathbb{Z},\ V_{j+1} \subset V_j$,
3. $\forall j \in \mathbb{Z},\ f(.) \in V_j \Leftrightarrow f(./2) \in V_{j+1}$,
4. $\lim_{j\to-\infty} V_j = \cap_{j=-\infty}^{j=+\infty} V_j = \{\emptyset\}$,
5. $\lim_{j\to+\infty} V_j = \mathbf{Closure}(\cup_{j=-\infty}^{j=+\infty} V_j) = \mathbf{L}^2(\mathbb{R})$,

In addition, for a sequence $\{V_j\}_{j\in\mathbb{Z}}$ to be a multi-resolution of $\mathbf{L}^2(\mathbb{R})$, there must exist a function θ in $\mathbf{L}^2(\mathbb{R})$ such that the family $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a basis of V_0.

Multi-resolution approximations have been extensively used in computer vision since their introduction in the article [16] of Burt and Adelson. From this perspective, a signal of dyadic size 2^J is the orthogonal projection of a function f in $\mathbf{L}^2(\mathbb{R})$ on some space $V_J \subset L^2(\mathbb{R})$. The approximation of the signal at a resolution 2^{-j}, with $j > J$, is defined as its orthogonal projection on a subspace V_j. In higher dimensions D, e.g. for images, multi-resolution approximations of $\mathbf{L}^2(\mathbb{R}^D)$ can be obtained by considering tensorial products between subspaces: $V_j^D = V_j \otimes V_j \otimes \ldots \otimes V_j$.

In our algorithm, we consider a multi-resolution approximation based upon a discrete image wavelet decomposition. It is possible to define a scaling function ϕ from the Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ of V_0. An approximation of the image at scale j is computed by projecting the approximation image at scale $j-1$ on the family $\{\phi_j(x-n)\}_{n\in\mathbb{Z}}$ of scaling functions, where

$$\phi_j(x) = \frac{1}{\sqrt{2^j}} \phi(\frac{x}{2^j}). \tag{3}$$

It can be shown that the projection can be computed by relying on iterative convolutions with a low-pass filter, followed by a factor 2 sub-sampling. The discrete decimated wavelet transform of the original image is obtained by iteratively filtering the approximation image by tensorial products of a low-pass filter and of a high-pass filter, yielding one approximation image and three details images at each iteration [15].

Fig. 1. Wavelet transform associated to Lena image for two decomposition levels. The size of the image is 512 by 512 pixels. The decomposition was performed using Daubechies wavelet with 12 vanishing moments. The approximation image corresponds to the subimage on the top left quarter. The other three subimages correspond to the projection of each approximation on a family of wavelets. We note that the main objects are relatively well preserved by each approximation image.

2.2 Hierarchical Contour Detection

A multi-resolution approximation of an image is presented in Fig. 1, for two levels of decomposition. The multi-resolution approximation is computed using Daubechies wavelets with 12 vanishing moments. Interestingly, we can note that the main objects and contours of the original image are relatively well preserved in its approximations. However, these images contain considerably less details than the original image, making them of potential interest for segmentation. Several methods have been proposed to handle segmentation using multi-resolution approaches. In 2000, Rezaee *et al.* [17] notably proposed an algorithm combining the pyramid transform and fuzzy clustering, obtaining good segmentation results on magnetic resonance images. In 2003, Kim and Kim [18] proposed a segmentation procedure relying on pyramidal representation and region merging.

It is straightforward to directly apply a watershed algorithm on the approximation images. However, the resolution of these images is significantly lower than the one of the original image. It is therefore difficult to establish a direct

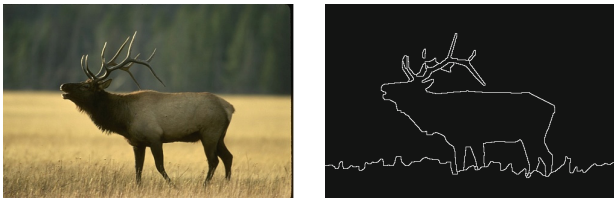

Fig. 2. Original image from the Berkeley segmentation database used to illustrate the algorithm, along with an example of human segmentation.

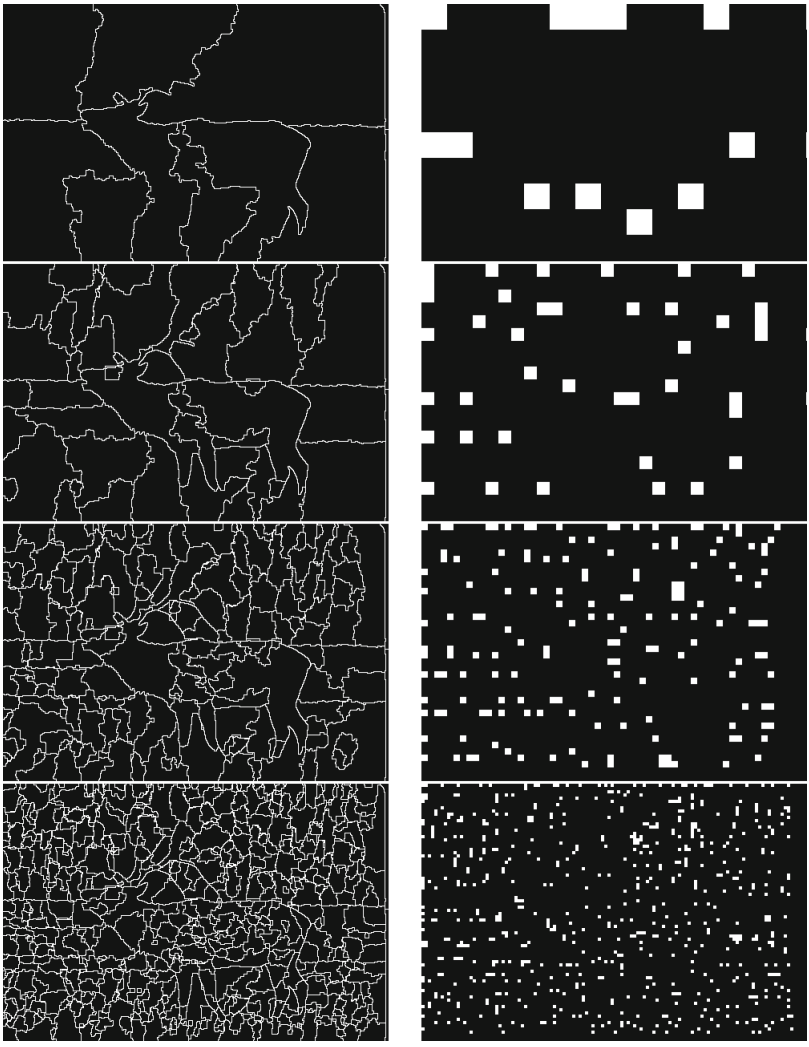

Fig. 3. Contour images corresponding to four decomposition levels (levels 4, 3, 2, 1 respectively) of the image presented in Fig. 2. The markers at each scale are displayed on the right images.

Fig. 4. Original image from the Berkeley segmentation database used to illustrate the algorithm, along with an example of human segmentation.

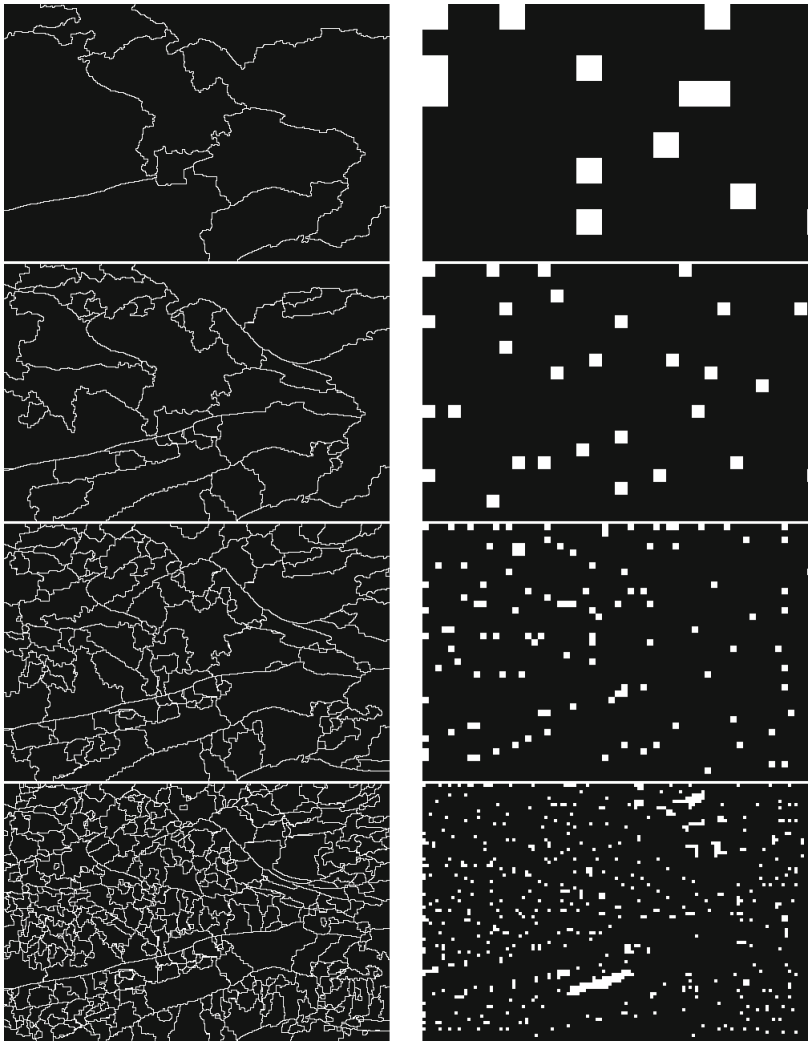

Fig. 5. Contour images corresponding to four decomposition levels (levels 4, 3, 2, 1 respectively) of the image presented in Fig. 4. The markers at each scale are displayed on the right images.

correspondence between the contours of the approximation image and the contours of the original image [18]. In this study, we propose to tackle this issue in a simple manner by defining multi-scale markers for performing the segmentation.

Let us consider an image I of dyadic size 2^J by 2^J pixels. We denote by L the number of decomposition levels. For l between 0 and L, we denote $I^{(l)}$ the approximation of I after l decomposition steps. The size of the image $I^{(l)}$ is therefore 2^{J-l} by 2^{J-l} pixels. By convention, $I^{(0)}$ is the original image I. $I^{(l+1)}$ is obtained by successively applying the low-pass filter associated to the wavelet transform to the rows and the columns of $I^{(l)}$ and down-sampling the result by a factor 2.

To initialize the contour detection algorithm, we first extract the minima of the gradient of the approximation image $I^{(L)}$. We obtain a sequence $\{m_1^{(L)}, ..., m_{K_L}^{(L)}\}$ of K_L markers. The locations of these markers are specified on an image $M^{(L)}$ of size 2^{J-L} by 2^{J-L}. To propagate the markers back to the original image I, we oversample the image $M^{(L)}$ by replacing each pixel $M^{(L)}(p,q)$ by an array of pixels of size 2^L by 2^L whose value is set equal to $M^{(L)}(p,q)$. Then, we apply the watershed algorithm on the image I from the image markers $M^{(L)}$, therefore obtaining a contour image $C^{(L)}$ associated to the approximation image $I^{(L)}$. We then repeat these operations at decomposition level $L-1$ to obtain a contour image $E^{(L-1)}$. We consider then the supremum image $S^{(L-1)}$ defined by

$$S^{(L-1)} = \sup(C^{(L)}, E^{(L-1)}), \tag{4}$$

where the supremum is defined as follows for each pixel $S^{(L-1)}[p,q]$:

$$S^{(L-1)}[p,q] = \max(C^{(L)}[p,q], E^{(L-1)}[p,q]). \tag{5}$$

Finally, we apply a watershed transform to the supremum image $S^{(L-1)}$ to obtain the contour image $C^{(L-1)}$ corresponding to decomposition level $L-1$. This last step is necessary to remove the thick contours that can potentially be created by the supremum between images $C^{(L)}$ and $E^{(L-1)}$. We obtain a multi-scale contour by iteratively applying the procedure described previously for all decomposition levels.

To illustrate the algorithm, the segmentation algorithm is applied to the images displayed in Figs. 2 and 4. The results are displayed in Figs. 3 and 5. We used Daubechies wavelets with 12 vanishing moments to calculate the successive multi-resolution approximations. By construction, the algorithm returns a nested sequence of contours, in the sense that if a pixel of the contour image $C^{(l)}$ is labelled as a contour, then the same pixel in the contour image $C^{(l-1)}$ is also labelled as a contour. The proposed approach therefore results in a hierarchical segmentation. We can see that as expected, the contours of the largest objects tend to be extracted for the approximations with the lowest resolution. By contrast, all contours corresponding to the details of the image to be segmented appear for the approximation images of higher resolution.

3 Experimental Results and Discussion

In Sect. 2, we introduced a hierarchical contour detection algorithm based upon successive multi-resolution approximations of the image to be segmented. It is of interest to assess the validity of the approach by comparing the results of the algorithm to human segmentations. To that end, we rely on the Berkeley Segmentation Database (BSD) [19]. The BSD is a large dataset of natural images that have been manually segmented. At each scale of the contours hierarchy, we compare the results of the detection algorithm to the human annotations.

In our comparison, we consider two criteria. We first estimate the proportion of contours detected by the algorithm and that have also been annotated (precision). To that end, we apply a dilation of size two to the contour image corresponding to the human segmentation, and we consider the intersection between the dilated image and the contour image returned by the algorithm at each decomposition level. The dilation is applied to account for potential inaccuracy of the human contour detection. The number of pixels belonging to the intersection normalized by the number of pixels corresponding to the detected contours provides us with an estimate of the proportion of detected contours that have also been manually segmented. The second criteria that we consider is the proportion of contours that have been detected by humans and that are also detected by the algorithm (recall). At each decomposition level, we apply a dilation of size two to the contour image returned by the algorithm and we consider the intersection between the dilated image and the contour image corresponding to the human segmentation. By counting the number of pixels of the intersection normalized by the number of pixels belonging to the human detected contours, we can determine the proportion of actual contours that are returned by the detection algorithm at each scale.

We estimated both criteria on 200 images from the BSD. The results are presented in Table 1. We can note that on average, the precision increases with the decomposition level in the contours hierarchy. This tends to assess the validity of the proposed multi-scale approach, in the sense that the contours detected with markers obtained after several decomposition levels have a significantly higher probability to correspond to human detected contours than contours directly obtained with the watershed transform. We also note that the recall of the algorithm decreases monotonously on average. This trend was to be expected, since the multi-scale approach inherently removes the contours of the smallest patterns. However, up to three decomposition levels, the recall remains higher than 0.5.

It is finally of interest to compare the results of the wavelet based markers selection to other commonly encountered methods, namely markers selection through h-minima of the image gradient and contour detection after filtering by alternate sequential filters. The difficulty here is to obtain a comparable number of segments between the distinct approaches. To that end, for each image, we select the value of h-minima and the size of structuring element for the alternate sequential filter, respectively, that yield the number of markers the closest from the one used in the wavelet based segmentation. We repeat the process for each decomposition

scale. Next, we rely on the aforementioned procedure to estimate the precision and the recall of these methods. The results are summarized in Table 1. We note that, on average, the segmentation based upon the discrete wavelet transform performs significantly better in terms of both precision and recall than the segmentation following an alternate sequential filtering. Our interpretation of this result is that the wavelet transform better preserves the contours of the original image at the highest scales of the transform. By contrast, for structuring elements of large size yielding a number of markers similar to the one obtained with the wavelet decomposition, alternate sequential filters significantly degrade the contours of the image, which explains the poor results that are registered in terms of precision and recall. Marker selection through h-minima values is shown to yield a higher precision than the wavelet based algorithm. However, in terms of recall, the wavelet based algorithm performs significantly better on average. Both approaches remain however significantly distinct and are difficult to compare, since a segmentation based upon markers selection by h-minima is highly sensitive to the local minima of the image gradient.

Table 1. Results of the wavelet based algorithm on the BSD for each decomposition level. Precision corresponds to the average proportion of contours detected by the algorithm that have also been annotated, along with the corresponding standard deviation. Recall corresponds to the average proportion of contours that have been detected by humans and that are also detected by the algorithm, along with the corresponding standard deviation. The proportions are obtained on a database of 200 images. The results of h-minima based segmentation and alternate sequential filtering based segmentation are also presented.

	Wavelet filtering		h-minima		Alternate sequ. filters	
Dec. level	Precision	Recall	Precision	Recall	Precision	Recall
0	0.10 ± 0.05	0.94 ± 0.03	0.12 ± 0.06	0.85 ± 0.16	0.10 ± 0.04	0.91 ± 0.06
1	0.13 ± 0.07	0.82 ± 0.07	0.17 ± 0.08	0.71 ± 0.18	0.12 ± 0.05	0.69 ± 0.09
2	0.17 ± 0.09	0.69 ± 0.11	0.24 ± 0.13	0.51 ± 0.14	0.12 ± 0.05	0.46 ± 0.08
3	0.22 ± 0.12	0.53 ± 0.14	0.30 ± 0.20	0.21 ± 0.08	0.12 ± 0.05	0.22 ± 0.07

4 Conclusion and Perspectives

In this article, we presented a new method to select markers for the watershed algorithm based upon multi-resolution approximations. By relying on the discrete decimated wavelet transform to obtain successive approximations of the image to be segmented, we were able to define a hierarchical structure for the detected contours. We evaluated the performance of the proposed approach by comparing its results to manually segmented images from the Berkeley segmentation database. The comparison provided an empirical evidence that the contours detected for the approximation of lowest resolutions have a higher probability to correspond to human detected contours than contours detected by the classical watershed transform.

An interesting perspective of this study is to use the proposed algorithm in a bottom-up aggregation procedure to obtain a dissimilarity measure between adjacent regions, corresponding to the decomposition scale at which the contour appears in the hierarchical segmentation. The empirical results obtained on the Berkeley Segmentation Database indicate indeed that regions separated by contours appearing early in the hierarchical segmentation procedure are less likely to be merged. We also noted on the comparison performed with the BSD that at the lowest resolutions, the recall of the detection method is around 0.5. A natural extension of this work is therefore to rely on additional statistical features that can eliminate wrong contours while keeping the actual contours in a bottom-up aggregation procedure.

References

1. Beucher, S., Lantuéjoul, C: Use of watersheds in contour detection (1979)
2. Meyer, F., Beucher, S.: Morphological segmentation. J. Vis. Commun. Image Represent. **1**(1), 21–46 (1990)
3. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. **13**(6), 583–598 (1991)
4. Meyer, F.: From connected operators to levelings. Comput. Imaging Vis. **12**, 191–198 (1998)
5. Meyer, F.: Levelings, image simplification filters for segmentation. J. Math. Imaging Vis. **20**(1–2), 59–72 (2004)
6. Jung, C.R., Scharcanski, J.: Robust watershed segmentation using wavelets. Image Vis. Comput. **23**(7), 661–669 (2005)
7. Jung, C.R.: Combining wavelets and watersheds for robust multiscale image segmentation. Image Vis. Comput. **25**(1), 24–33 (2007)
8. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer Science & Business Media, Berlin (2013)
9. Cheng, J., Rajapakse, J.C., et al.: Segmentation of clustered nuclei with shape markers and marking function. IEEE Trans. Biomed. Eng. **56**(3), 741–748 (2009)
10. Angulo, J., Jeulin, D.: Stochastic watershed segmentation. In: Proceedings of the 8th International Symposium on Mathematical Morphology, pp. 265–276 (2007)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. Int. J. Comput. Vis. **59**(2), 167–181 (2004)
12. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 315–327 (2012)
13. Beucher, S.: Watershed, hierarchical segmentation and waterfall algorithm. In: Serra, J., Soille, P. (eds.) Mathematical Morphology and its Applications to Image Processing, pp. 69–76. Springer, Heidelberg (1994)
14. Marcotegui, B., Beucher, S.: Fast implementation of waterfall based on graphs. In: Ronse, C., Najman, L., Decencière, E. (eds.) Mathematical Morphology: 40 Years On, pp. 177–186. Springer, Heidelberg (2005)
15. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press, Cambridge (1999)
16. Burt, P., Adelson, E.: The Laplacian pyramid as a compact image code. IEEE Trans. Commun. **31**(4), 532–540 (1983)

17. Rezaee, M.R., Van der Zwet, P.M.J., Lelieveldt, B.P.E., Van Der Geest, R.J., Reiber, J.H.C.: A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. IEEE Trans. Image Process. **9**(7), 1238–1248 (2000)
18. Kim, J.-B., Kim, H.-J.: Multiresolution-based watersheds for efficient image segmentation. Pattern Recogn. Lett. **24**(1), 473–488 (2003)
19. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV, vol. 2, pp. 416–423. IEEE (2001)