Efficient Optimal Multi-level Thresholding for Biofilm Image Segmentation

Darío Rojas[1], Luis Rueda[2], Homero Urrutia[3], and Alioune Ngom[2]

[1] Department of Computer Science, University of Atacama, 485 Copayapu Ave.,
Copiapó, Chile
`dario.rojas@uda.cl`

[2] School of Computer Science, University of Windsor, 401 Sunset Ave., Windsor,
ON, N9B 3P4, Canada
`{lrueda,angom}@cs.uwindsor.ca`

[3] Biotechnology Center and Faculty of Biological Sciences, University of Concepción,
Barrio Universitario, Concepción, Chile
`hurrutia@udec.cl`

Abstract. A microbial biofilm is structured mainly by a protective sticky matrix of extracellular polymeric substances. The appreciation of such structures is useful for the microbiologist and can be subjective to the observer. Thus, quantifying the underlying images in useful parameters by means of an objective image segmentation process helps substantially to reduce errors in quantification. This paper proposes an approach to segmentation of biofilm images using *optimal* multilevel thresholding and indices of clustering validity. A comparison of automatically segmented images with manual segmentation is done through different thresholding criteria, and clustering validity indices are used to find the correct number of thresholds, obtaining results similar to the segmentation done by an expert.

1 Introduction

A biofilm is a complex aggregation of microorganisms that live upon surfaces, structured mainly by the secretion of a protective sticky matrix of extracellular polymeric substances. In order to understand biofilm structures, scanning electronic microscopy (SEM), confocal laser scanning microscopy (CLSM), and optical microscopy (OM) are currently used [1]. The appreciation of such structures in digital images can be subjective to the observer [2], and hence it is necessary to quantify the underlying images in useful parameters for the microbiologist. Automatic segmentation is crucial in this regard, which, if done in correct form, it does not propagate errors of appreciation in image quantification. However, the evaluation of automatic segmentation algorithms is subjective, leaving to the designer the responsibility to judge the effectiveness of the technique based only on their intuition and results of some examples of image segmentation. To solve this problem, the work presented in [3] demonstrated the effectiveness of the normalized probabilistic Rand's Index (RI), which can be used to make a

quantitative comparison between different algorithms for segmentation by means of a set of manually segmented images. The RI index was first presented in [4], and is based on the original work of [5].

On the other hand, in [6,7], two novel methods called COMSAT and PHLIP were proposed, which are able to quantify the characteristics of biofilms obtained through CLSM image stacks. They both use Otsu's thresholding criterion [8] for image segmentation, but no further studies have been performed regarding the efficiency of the segmentation. Another work related to quantifying the parameters of biofilm structures was presented in [2], and a detailed explanation of the quantification methods can be found in [9]. The algorithms for segmentation used by these approaches are also based on Otsu's criterion and an iterative method for finding thresholds that was proposed in [2]. In the same context, in [10], a review of several automatic thresholding algorithms for biofilm segmentation was presented, including local entropy, joint entropy, relative entropy, Renyi's entropy and iterative selection, but none of these methods were used to evaluate criteria for multi-level thresholding. Furthermore, the method for evaluating the differences between segmented and original images is the sum of squares of relative residuals (MSSRR). This method takes threshold values into account but not the differences between images. In addition, the approach of [10] is not capable of comparing the results of images segmented with multi-level thresholding algorithms or images segmented with different numbers of thresholds.

All approaches for automatic image segmentation based on thresholding proposed so far do not allow to segment different kinds of biofilm images optimally and without the intervention of an expert. In this paper, an approach to segmentation of biofilm images is proposed, based on optimal multi-level thresholding, which is carried out in *polynomial time*. Also, clustering validity indices are used for finding the best number of thresholds automatically.

2   The Proposed Method

A method for segmentation of biofilm images was implemented through an efficient optimal multi-level thresholding algorithm. Different thresholding criteria and clustering validity indices were implemented for measuring the performance of segmentation methods and the determination of the best number of thresholds, respectively.

2.1   Polynomial-Time Optimal Multi-level Thresholding

In [11], a polynomial-time algorithm for multi-level thresholding was proposed. This algorithm is polynomial not just on the number of bins of the histogram, but also on the number of thresholds. Moreover, it runs in polynomial time independently of the thresholding criterion. In [11], we defined the optimal solution by searching an optimal set of thresholds, $T = \{t_0, \ldots, t_k\}$, that maximizes a function Ψ as follows:

$$\Psi(T) = \sum_{j=1}^{k+1} \psi_{t_{j-1}+1, t_j} \tag{1}$$

where $\Psi : P^k \times [0,1]^n \mapsto \mathbb{R}^+$, k is the number of thresholds, $P = \{p_1, \ldots, p_n\}$ are the probabilities of the bins in the histogram, n is the number of bins in the histogram, and function $\psi_{t_{j-1}+1,t_j} : P^2 \times [0,1]^{t_j-t_{j-1}+2} \mapsto \mathbb{R}^+ \cup \{0\}$ (where t_j is the jth threshold of T) must satisfy the following conditions:

1. For any histogram P and any threshold set T, $\Psi > 0$ and $\psi \geq 0$.
2. For any m, $1 \leq m \leq k+1$, $\Psi(\{t_0, ..., t_m\})$ can be expressed as $\Psi(\{t_0, ..., t_{m-1}\}) + \psi_{t_{m-1}+1,t_m}$.
3. If $\psi_{t_{j-1}+1,t_j}$ is known, then $\psi_{t_{j-1}+2,t_j}$ can be computed in $O(1)$ time.

The three thresholding criteria are defined as follows:

$$\text{Otsu's (OTSU)} : \psi_{t_{j-1}+1,t_j} = \omega_j \mu_j^2, \tag{2}$$

$$\text{Minimum Error (MINERROR)} : \psi_{t_{j-1}+1,t_j} = 2\omega_j \{\log \sigma_j + log\omega_j\}, \tag{3}$$

$$\text{Entropy-based (ENTROPY)} : \psi_{t_{j-1}+1,t_j} = - \sum_{i=t_{j-1}+1}^{t_j} \frac{p(i)}{\omega_j} log \frac{p(i)}{\omega_j}, \tag{4}$$

where t_j is the jth threshold of T, $\omega_j = \sum_{i=t_{j-1}+1}^{t_j} p(i)$, $\mu_j = \frac{1}{\omega_j} \sum_{t=t_{j-1}+1}^{t_j} ip(i)$ and $\sigma_j = \frac{1}{\omega_j} \sum_{i=t_{j-1}+1}^{t_j} h_p(i)(i - \mu_j)^2$.

It is important to note that biofilm images lead to *sparse* histograms (many bins have zero probabilities), and so, for the sake of efficiency the algorithm for *irregularly* sampled histograms as presented in [11] is implemented in our work.

2.2   Optimal Number of Thresholds

The optimal thresholding algorithm discussed above is not able to determine the number of thresholds, k, in which the image can be segmented correctly. However, k has a direct relationship with the number of classes, $k+1$, in which a histogram is partitioned. Viewing thresholding as a problem of clustering pixel intensities, clustering validity indices can be used to obtain the best number of classes, $k+1$, and hence the number of thresholds. In this work, the Davies-Bouldin Index (DB) is used, which is defined as the ratio between the within-cluster scatter versus the between-cluster scatter. The goal is to minimize the value of the DB function, defined as [12]:

$$\text{DB} = \frac{1}{k+1} \sum_{i=1}^{k+1} max_{1 \leq j \leq k+1, j \neq i} \left\{ \frac{S_i + S_j}{d_{ij}} \right\}, \tag{5}$$

where $k+1$ is the number of clusters, $S_j = \frac{1}{|\zeta_j|} \sum_{i=t_{j-1}+1}^{t_j} p(i)||i - \mu_j||$ is the within-cluster scatter of cluster ζ_j, $d_{ij} = ||\mu_i - \mu_j||$ is the distance between clusters ζ_i and ζ_j. Other validity indices [12] such as Dunn's index (DN), Calinski-Harabasz's index (CH) and index I (IndexI) were also evaluated to compare the results.

2.3   Manual Segmentation vs Automatic Segmentation

As discussed previously, biofilm structures contain information about phenotypic effects of bacteria. Many studies [13,14,15,16] indicate that the structural heterogeneity of a biofilm can affect its dynamic activity and functional properties. In order to obtain this structural information, the segmentation of images in an unsupervised and optimal form (for an objective comparison and reproducibility of experiments) is an important component in any image analysis model for biofilm quantification.

In this context, some clustering algorithms used for segmentation do not guarantee an optimal solution, because they require the specification of good initial cluster centers for correct convergence and the solution can be a local optimum [17]. Segmentation approaches such as region growing methods are more tolerant to noise, however, they require the specification of an appropriate initial set of seeds to produce accurate results [18]. Segmentation methods such as image filtering may miss fine but important details of confocal images and hence, are only recommended for image visualization, not for image quantification [2].

To determine the best thresholding criterion, a similarity index of partitions is used, namely the *Probabilistic Rand Index* (RI), which is the percentage of pairs for which there is an agreement. Let $L = \{l_1, ..., l_N\}$ and $L' = \{l'_1, ..., l'_N\}$ be the ordered sets of labels l_i and l'_i respectively, for each pixel $1 \leq i \leq N$ of two segmented images to be compared, the RI index is defined as follows:

$$\text{RI}(L, L') = \frac{1}{\binom{N}{2}} \sum_{i,j \ i \neq j}^{N} [I(l_i = l_j) \wedge I(l'_i = l'_j) + I(l_i \neq l_j) \wedge I(l'_i \neq l'_j)] \qquad (6)$$

where I is the identity function and $\binom{N}{2}$ is the total number of pairs among the N pixels. This index takes a value one when L and L' are equal, and zero if they do not agree on anything at all. The best technique for automatic segmentation of biofilm images was found experimentally, using a method that combines automatic multi-level thresholding and clustering validity index is proposed. Biofilm images were acquired from optical and confocal microscopy, and the histogram for each image was computed. Manual multi-level segmentation was performed by an expert, by means of a trial and error process, in order to determine the best number of thresholds, k, and set of thresholds, T. Furthermore, each original image was automatically segmented, by means of the optimal multi-level thresholding algorithm proposed in [11], using the three thresholding criteria mentioned above, and for several values of k. Finally, clustering validity indices were computed for each image that was segmented automatically and the RI index was calculated for each segmented image, by means of manual and automatic thresholding, in order to determine the best thresholding criteria.

3   Experimental Results

A dataset of 649 images was used to perform the evaluation of the proposed biofilm segmentation approach. These images were obtained as follows. Mature

Table 1. Dataset for determining the best combination of techniques

k	No. of images	Microscopy	Resolution
1	616	Confocal	512×512
2	10	Optical	1040×1392
3	10	Optical	1040×1392
4	6	Optical	1040×1392
5	6	Optical	1040×1392
6	1	Optical	1040×1392

biofilms of *Pseudomonas syringae* strains were developed within the Biofilm and Environmental Microbiology Laboratory[1]. These biofilms were then scanned using CLSM and OM, generating stacks of images that represent the three-dimensional structure of the biofilms. The images were segmented individually, manually and automatically. Table 1 shows the features of each image (images are in 12-bit grayscale) with its number of thresholds, k, found manually by an expert.

In order to avoid any bias introduced by the differences in the nature of the images, the dataset was divided into two subsets of images. The set of images which are best segmented with one threshold (obtained from CLSM), and the set of images which are best segmented with more than one threshold (obtained from OM).

3.1 Performance of Thresholding Criteria

The best thresholding criterion was found by using the RI index comparing manual vs. automated segmentation, and by using the following notation: RI^{all} is the RI index for all image datasets, RI^{clsm} is the RI index for images with one threshold found manually, and RI^{op} is the RI index for images with more than one threshold found manually. In Table 2, the resulting values for the RI index are depicted for all image subsets. It is clear that ENTROPY is the best criterion for segmenting images with one threshold. On the other hand, OTSU is the best criterion for segmentation of images with more than one threshold. Overall, the ENTROPY criterion achieved the best performance for all image datasets.

Table 2. RI index for different image subsets (number of thresholds found manually)

Index and Dataset	OTSU	ENTROPY	MINERROR
RI^{op}	**0.7897**	0.7300	0.7713
RI^{clsm}	0.7283	**0.7767**	0.6086
RI^{all}	0.6184	**0.7566**	0.5846

---

[1] http://www.udec.cl/~bem-lab/

Table 3. MSSRR and R^2 for different thresholding criteria (CLSM images)

	OTSU	ENTROPY	MINERROR
MSSRR	3.851	**0.827**	10.25
R^2	0.71	**0.76**	0.62

Additionally, for images obtained from CLSM, the mean of the sum of squared relative residuals (MSSRR) and correlation R^2 were calculated in order to evaluate the differences between the threshold levels selected by automatic thresholding and the manual method. The MSSRR is defined as follows [10]:

$$\text{MSSRR} = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{t_i - t_i'}{t_i} \right)^2 \tag{7}$$

where t_i and t_i' are the ith thresholds for images found manually and automatically, respectively, and $M = 616$ is the total number of images from CLSM. Figs. 1(a)(b)(c) show correlation plots between manually-obtained sets of thresholds and thresholds obtained from ENTROPY, MINERROR and OTSU, respectively. For an agreement between automatic and manual thresholding we expect that data points follow the diagonal line $(y = x)$. In Table 3, the resulting values for MSSRR and R^2 for each thresholding criterion are shown.

3.2   Determining the Best Number of Thresholds

The number of thresholds has a direct relation with the number of classes in which an image can be segmented. Therefore, a separate measurement on the error was used to estimate the number of thresholds for the complete dataset of images. Table 4 shows the Mean Squared Error (MSE) for each combination of thresholding criteria and clustering validity indices. As can be observed, the DB index achieves the best performance before applying the ENTROPY thresholding criterion, which reaffirms that the combination ENTROPY+DB attains the best performance in most of the cases for different datasets of images.

The clustering validity indices have a direct relationship among them in their formulation [12]; however, each index has a different behavior depending on the number of thresholds selected. The behavior of each validity index can be seen in

Table 4. Mean Squared Error (MSE) for the estimation of the best number of thresholds in all datasets

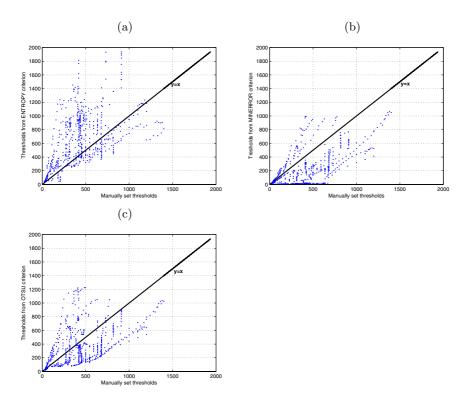|	IndexI	CH	DB	DN
OTSU	7.44	221.97	96.63	212.76
ENTROPY	2.33	212.2	**1.18**	186.94
MINERROR	2.80	188.31	179.39	220.22

Fig. 1. Correlation plots between thresholds obtained manually and automatically using different criteria: (a) ENTROPY, (b) MINERROR, and (c) OTSU

Fig. 2. Although the plots are for one of the images in the dataset, they represent the general behavior of the clustering validity indices for an entire dataset.

In Fig. 2, we notice that the indices *IndexI, CH* and *DN* are (for most of the values of k) monotonically increasing functions of k (Figs. 2 (a) (b) (d)), obtaining the best performance when the function achieves the maximum, i.e. when $k = 61$, $k = 64$ and $k = 64$ respectively. This behavior, unfortunately, does not provide a clear direction on how to determine what is the optimal number of clusters with which an image should be segmented. It also illustrates the high MSE values obtained by these indices to estimate the best number of thresholds. On the other hand, the DB index is the only index that shows a high independence in terms of the number of clusters. This index reaches its optimal performance when $k = 8$, which is a much more meaningful value than those obtained by the others. Moreover, as k grows, the DB index tends to give an almost constant rate, which reflects the fact that increasing the number of clusters, the quality of the clustering does not improve from a certain point (Fig. 2 (c)).
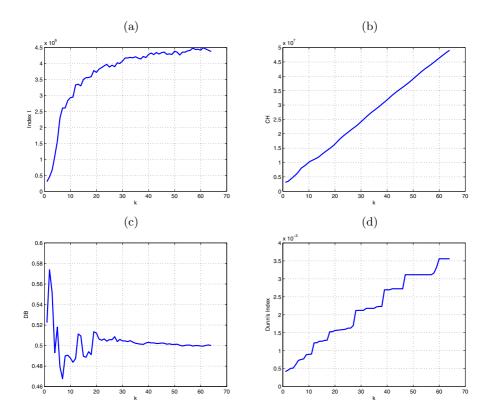
Fig. 2. General behavior of clustering validity indices: (a) IndexI, (b) CH, (c) DB, (d) DN

3.3 Performance for Image Segmentation Techniques Combined with Cluster Validdty Indices

Table 5 shows the performances of the RI index for all biofilm images. The table shows that the best combination is ENTROPY + DB for RI^{all} index. Also, it is clear that the thresholding criteria with the best performance for this dataset is based on the ENTROPY criterion. This result was predictable, because most of the images have one threshold and the best method of segmentation for one threshold is the ENTROPY criterion.

Table 5. The RI[all] index for all automatically segmented biofilm images

	IndexI	CH	DB	DN
OTSU	0.2163	0.2151	0.2969	0.2187
ENTROPY	0.2506	0.2351	**0.7884**	0.2385
MINERROR	0.2206	0.2332	0.2613	0.2085

The behaviors for different combinations of techniques for two separate cases, one threshold and more than one threshold, are discussed next.

3.4 One Threshold

All biofilm images obtained by the confocal microscopy have a single optimal threshold (manually found by the expert). Table 6 shows the performance of thresholding criteria and clustering validity indices for the segmentation of biofilm images with one threshold determined automatically.

Table 6. The RI^{clsm} index for automatically segmented images of biofilms with one threshold determined automatically

	IndexI	CH	DB	DN
OTSU	**0.6176**	0.3901	0.5297	0.4002
ENTROPY	0.7573	0.4907	**0.7634**	0.5029
MINERROR	**0.5844**	0.3279	0.328	0.3075

The best performance is reached by the combinations of methods ENTROPY and DB, corroborating the overall results. In this case, the analysis shows the same pattern as that of the overall performance, because the ENTROPY criterion is the best criterion for segmenting images with one threshold and the DB index is the best clustering validity index for an estimated value of k.

3.5 More Than One Threshold

Table 7 shows the performance of thresholding methods and cluster validity indices for segmentation of biofilms with more than one threshold. As can be seen, all methods achieve a very good performance. The OTSU criterion combined with IndexI attains the best value of the RI^{op} index, but for this set of images, the performance with respect to the combination ENTROPY+DB do not differ significantly. However, it is clear that the performances of thresholding criteria are significantly influenced by the number of classes estimated by the clustering validity indices.

Table 7. The RI^{op} index for automatically segmented biofilm images with more than one threshold determined automatically

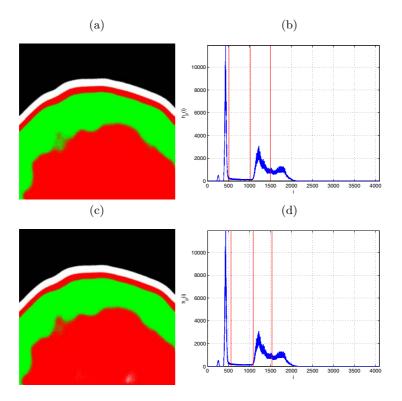|	IndexI	CH	DB	DN
OTSU	**0.7739**	0.6548	0.7070	0.6564
ENTROPY	0.6889	0.7046	**0.7634**	0.7077
MINERROR	**0.7594**	0.6657	0.7222	0.6302

Fig. 3. Multi-level Thresholding Segmentation: (a) Optical image segmented manually. (b) Histogram of (a). (c) Optical image segmented automatically. (d) Histogram of (c).

3.6   Visual Validation

Figs. 3(a) and (c) show the manual segmentation of a biofilm image with more than one threshold compared to the automatic segmentation that combines OTSU + IndexI. As can be seen, in Fig. 3(c), the result of automatic segmentation is close to that of the manual segmentation (Fig. 3(a)), setting the thresholds to almost the same values when the segmentation is done by an expert (Figs. 3(b) and (d)).

On the other hand, rebuilding the structure of a biofilm from images of CLSM, offers a powerful visualization tool. Figs. 4(a) and (b) show the 3D reconstruction of a biofilm through images segmented manually, and automatic biofilm reconstruction by means of images segmented automatically through combinations of techniques ENTROPY+DB detecting optimal thresholds. As can be seen, the image reconstructed automatically, is quite similar to a manual reconstruction done by an expert.
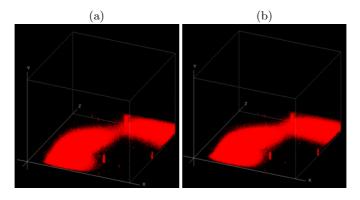
Fig. 4. Biofilm reconstruction: (a) Manual reconstruction. (b) Automatic reconstruction.

4 Conclusions

A method for automatic segmentation of biofilm images has been proposed. This method is based on the entropy-based criterion for multi-level thresholding, estimating the number of thresholds through the well-known Davies-Bouldin clustering validity index. This index is able to find the best number of thresholds close to the criteria established by an expert. This was assessed by using objective measures, the *Probabilistic Rand Index*, which compares results between segmentations done by the proposed method and the segmentation done by an expert. Automatic segmentation of biofilm images leads to a much better quantification process and also to better understanding of bacterial biofilms. Since the multi-level thresholding used is always optimal, it is possible to make a segmentation process free of subjectivity. Although the three main thresholding criteria have been implemented and tested in this work, other criteria can also be used for the segmentation of biofilm images, whenever they satisfy the conditions stated in [11], and for which optimal thresholding can be achieved in polynomial time.

References

1. Claxton, N.S., Fellers, T.J., Davidson, M.W.: Laser scanning confocal microscopy. Technical report, Department of Optical Microscopy and Digital Imaging, The Florida State University (2006)
2. Beyenal, H., Donovan, C., Lewandowski, Z., Harkin, G.: Three-dimensional biofilm structure quantification. Journal of Microbiological Methods 59, 395–413 (2004)

3. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 929–944 (2007)
4. Unnikrishnan, R., Hebert, M.: Measures of similarity. In: Proceedings of Seventh IEEE Workshop on Applications of Computer Vision, Application of Computer Vision, vol. 1, pp. 394–401 (2005)
5. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66, 846–850 (1971)
6. Heydorn, A., Nielsen, A.T., Hentzer, M., Sternberg, C., Givskov, M., Ersboll, B.K., Molin, S.: Quantification of biofilm structures by the novel computer program comstat. Microbiology 146, 2395–2407 (2000)
7. Mueller, L.N., de Brouwer, J.F., Stal, J.S.A., Xavier, L.J., Xavier, J.B.: Analysis of a marine phototrophic biofilm by confocal laser scanning microscopy using the new image quantification software phlip. BMC Ecology 6, 1–15 (2006)
8. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9, 62–66 (1979)
9. Beyenal, H., Lewandowski, Z., Harkin, G.: Quantifying biofilm structure: Facts and fiction. Biofouling 20, 1–23 (2004)
10. Yang, X., Beyenal, H., Harkin, G., Lewandowsi, Z.: Evaluation of biofilm image thresholding methods. Water Sci. Technology 35, 1149–1158 (2001)
11. Rueda, L.: An efficient algorithm for optimal multilevel thresholding of irregularly sampled histograms. In: da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C., Loog, M. (eds.) S+SSPR 2008. LNCS, vol. 5342, pp. 602–611. Springer, Heidelberg (2008)
12. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 1650–1655 (2002)
13. Costerton, J.W., Douglas, Z.L., Caldwell, E., Lappin-Scott, H.M., Korber, D.R.: Microbial biofilms. Annual Review of Microbiology 49, 711–745 (1995)
14. Johnson, L.R.: Microcolony and biofilm formation as a survival strategy for bacteria. Journal of Theoretical Biology 251, 24–34 (2008)
15. Jorgensen, T.M., Haagensen, J., Sternberg, C., Molin, S.: Quantification of biofilm structure from confocal imaging. Technical report, Optics and Fluids Department, Riso National Laboratory (2003)
16. Klapper, I.: Effect of heterogeneous structure in mechanically unstressed biofilms on overall growth. Bulletin of Mathematical Biology, 809–824 (2006)
17. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press, London (2003)
18. Adams, R., Bischof, L.: Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 641–647 (1994)