Image segmentation of biofilm structures using optimal multi-level thresholding

Article *in* International Journal of Data Mining and Bioinformatics · May 2011

DOI: 10.1504/IJDMB.2011.040384 · Source: PubMed

5 authors, including:

Dario Rojas
University of Concepción
35 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Luis Rueda
University of Windsor
182 PUBLICATIONS **1,031** CITATIONS
SEE PROFILE

Alioune Ngom
University of Windsor
141 PUBLICATIONS **1,037** CITATIONS
SEE PROFILE

Homero Urrutia
University of Concepción
44 PUBLICATIONS **438** CITATIONS
SEE PROFILE

Image segmentation of biofilm structures using optimal multi-level thresholding

Darío Rojas

Department of Computer Science,
University of Atacama,
485 Copayapu Ave.,
Copiapó 1532296, Chile
E-mail: dario.rojas@uda.cl

Luis Rueda* and Alioune Ngom

School of Computer Science,
University of Windsor,
401 Sunset Ave.,
Windsor, ON N9B 3P4, Canada
E-mail: lrueda@uwindsor.ca
E-mail: angom@uwindsor.ca
*Corresponding author

Homero Hurrutia and Gerardo Cárcamo

Center for Biotechnology
and Faculty of Biological Sciences,
University of Concepción, 4070386, Chile
E-mail: hurrutia@udec.cl
E-mail: gecarcamo@uder.cl

Abstract: The appreciation of biofilm structures in digital images can be subjective to the observer, and hence it is necessary to analyse the underlying images in useful parameters by means of quantification that is, ideally, free of errors. This paper proposes a combination of techniques for segmentation of biofilm images through an optimal multi-level thresholding algorithm and a set of clustering validity indices, including the determination of the best number of thresholds. The results, which are validated through Rand Index and a quantification process performed in a laboratory, are similar to the quantification and segmentation done by an expert.

Keywords: clustering; biofilm image processing; image segmentation; multi-level thresholding; biofilm quantification; biofilm structures.

Reference to this paper should be made as follows: Rojas, D., Rueda, L., Ngom, A., Urrutia, H. and Cárcamo, G. (2011) 'Image segmentation of biofilm structures using optimal multi-level thresholding', *Int. J. Data Mining and Bioinformatics*, Vol. 5, No. 3, pp.266–286.

Biographical notes: Darío Rojas received his Bachelor's Degree in Computer Science from the University of Concepción, Chile, in 2006 and his Master's Degree in Computer Science from the University of Concepción, in 2008, under the supervision of Luis Rueda. He is currently a Lecturer in the Department of Computer Science, University of Atacama, Chile. His research interests include issues related to pattern recognition, bioinformatics and multi-agent systems.

Luis Rueda received his Bachelor's in Computer Science from the National University of San Juan, Argentina, in 1993, and his Master's and PhD Degrees in Computer Science from Carleton University, Canada, in 1998 and 2002, respectively. He has been with the School of Computer Science, University of Windsor, Canada, as an Assistant Professor, from 2002 to 2005, and since then an Associate Professor. His main research interests are in pattern recognition, image processing, proteomics and microarrays. He holds one patent and more than 70 publications in prestigious journals and conferences. He has won various awards for university research achievements and conferences. He is a Senior Member of the IEEE and a Member of the IAPR.

Alioune Ngom is an Associate Professor at the University of Windsor, Ontario, Canada. Prior to joining the University of Windsor, he held the position of an Assistant Professor at the Department of Mathematics and Computer Science at Lakehead University (Thunder Bay, Ontario, Canada), 1998–2000. During his short stay at Lakehead University, he co-founded Genesis Genomics Inc. in 1999, a biotechnology company specialising in the analysis of mitochondrial genome and the design of biomarkers for the early detection of cancer. His main research interests include, but are not limited to, computational intelligence and machine learning methods and their applications in computational biology and bioinformatics problems such as: microarray analysis, protein analysis, oligonucleotide selection, bio-image analysis, and gene regulatory network analysis.

Homero Urrutia received his Bachelor's Degree in Biology from the University of Concepción, Chile, in 1982. He obtained his Master's Degree in Microbiology from the University of Concepción in 1992 and his PhD, in 1993, on Environmental Sciences in the context of the Europe–Latin America collaboration program (EULA–University of Concepción). He is currently an Associated Professor at the Faculty of Biological Science and a Senior Researcher at the Biotechnology Center, University of Concepción. His research is focused on the structure and biotechnological applications of microbial biofilms.

Gerardo Cárcamo received his Bachelor's Degree in Biochemistry and Master's in Microbiology at the University of Concepción, Chile, in 2007 and 2009 respectively. He is currently pursuing his PhD in Microbiology and Immunology at the University of Zurich, Switzerland. His research focuses on the involvement of quorum sensing and bacterial motility in biofilm formation.

1 Introduction

For several years, it has been shown that bacteria can form some types of societies by means of emergent behaviour to develop complex tasks impossible to be carried out by themselves individually according to Johnson (2008). One of the most remarkable emergent behaviours of bacteria is the formation of biofilms. A biofilm is a complex aggregation of microorganisms, structured mainly by the secretion of a protective sticky matrix of Extracellular Polymeric Substances (EPS). Also, according to Nikolaev and Plakunov (2007), a biofilm is commonly characterised by its adherence to the surfaces, structural heterogeneity, genetic diversity and complex community interactions. A Biofilm develops when free-living bacteria sense some environmental conditions (decrease or increase of nutrient availability, oxygen, temperature, etc.), which change their state from the free-living (planktonic) state to a growing biofilm state. The environment of the biofilm can vary from one organism to another. Bacteria can colonise biotics and abiotics surfaces, and environments inhabited by superior forms of life and a great variety of places inhospitable for most organisms. Their survival ability is explained partly by the metabolic versatility and phenotype plasticity, as indicated in a study conducted by Johnson (2008). The bacterial behaviour in a biofilm state is different from its planktonic counterpart as highlighted in Stewart and Franklin (2008), and can produce negative consequences in health, clinical, agricultural, and industrial environments according to Beech et al. (2005) and Hall-Stoodley et al. (2004), depending on the bacterial species. Also, bacteria in a biofilm state are extremely resistant to stresses in their environment such as antibiotics. The biological success of biofilms is not trivial as they involve several factors such as the use of genetic information, food and energy supply, which have profound consequences for bacterial physiology and survival. Medical and environmental scientists and microbiologists are interested in methods that address the negative effects of biofilms in given environments, including biofilm control approaches. The success of biofilm control methods depends on the ability to block the formation of the biofilm or inhibit its growth at given stages of its evolution. The analysis of biofilm images is among the means used by these scientists for understanding the dynamics of bacterial biofilms and their spatio-temporal evolutions in growing biofilms.

In order to understand biofilm structures, nowadays, special microscopies are used to obtain digital images from live bacterial biofilm structures, including Confocal Laser Scanning Microscopy (CLSM) and Optical Microscopy (OM). The appreciation for such structures in digital images can be subjective to the observer according to Beyenal et al. (2004a), and hence it is necessary to quantify the underlying images in useful parameters for the microbiologist. In order to quantify these images, pre-processing is necessary to distinguish relevant elements and characteristics from the structures. This involves an image segmentation process, which, if done in correct form, it does not propagate errors of appreciation in image quantification. The segmentation method discussed in this paper, involves determining an optimal set of thresholds such that pixels are correctly assigned to one of the groups. Biofilm structures represent information about phenotype effects of bacteria. Studies carried out by Johnson (2008), Costerton et al. (1995), Jorgensen et al. (2003) and Klapper (2006) indicate that the structural heterogeneity of a biofilm can affect its dynamic activities and functional propierties. Yang et al. (2000)

proposed a biofilm image processing approach, which aims to obtain structural characteristics. In the same work, an extension to the works of Beyenal et al. (1998), Lewandowski et al. (1992), Yang and Lewandowski (1995) was proposed, incorporating the computation of parameters like porosity, fractal dimension, diffusion capacity and entropy.

On the other hand, Heydorn et al. (2000) proposed COMSTAT, a software package that is able to quantify the characteristics of biofilms obtained through CLSM image stacks. However, this software uses manual thresholding methods for image segmentation. Another software for biofilm image segmentation was proposed by Tolle et al. (2003) is MAPPER, which has also been used to quantify biofilm images. However, the segmentation methods they used are manual. Another work related to quantifying the parameters of biofilm structures was carried out by Beyenal et al. (2004a). The algorithms for segmentation used by these approaches are the traditional criterion proposed by Otsu (1979), and an iterative method for finding the thresholds that was proposed by Beyenal et al. (2004a). Many software packages can now automatically determine the thresholds for image segmentation, and also, an increasing number of studies use CLSM for biofilm visualisation. However, Merod et al. (2007) found that certain CLSM image stacks can contain errors such as strange pixels which can lead to finding erroneous thresholds for the segmentation process. In order to resolve this problem, Merod et al. (2007) proposed a software called PHLIP, which makes an automatic image exclusion of the stack.

All approaches for automatic image segmentation based on thresholding proposed so far do not allow to segment different kinds of biofilms in an optimal way without the intervention of an expert. Since they need to set parameters manually, they produce sub-optimal solutions, subjective results, and they are prone to inefficient use of computational resources, or they do not find the number of thresholds automatically. In this paper, a combination of multi-level thresholding criteria and a set of clustering validity indices is proposed to perform a complete automatic and efficient optimal multi-level thresholding segmentation of different kinds of biofilm images.

2 The segmentation process

An efficient method for segmentation of biofilm images is proposed through a multi-level thresholding algorithm in an automatic form. Different criteria and clustering validity indices are implemented for measuring the performance of segmentation methods and the determination of the best number of thresholds respectively. Finally, we compare the resulting image segmentation against a set of images segmented manually by an expert.

2.1 The thresholding problem

An image, can be considered as a two dimensional discrete function of two independent spatial variables. A greyscale image, which is composed only of pixel intensities can be represented as a two dimensional function $f : \mathbb{N}^2 \to \mathbb{N}$, where $1 \leq f(x, y) \leq n$, n is the number of different intensity values, (x, y) with $1 \leq x \leq N$ and $1 \leq y \leq M$, are independent spatial variables indicating the position of the pixel in a 2D image, and N and M are the width and height of the image, respectively.

A frequency histogram H_f of an image with discrete intensity values into $[0, 2^n - 1]$, is an ordered set $H_f = \{h_f(1), h_f(2), \ldots, h_f(n)\}$, where $h_f(i)$ is a discrete intensity function that represents the frequency of pixels with intensity i in the image. Thus, the frequency $h_f(i)$ of the i^{th} intensity is $h_f(i) = \sum_{x=1}^{N} \sum_{y=1}^{M} (f(x, y) = i)$. Also, the histogram can be defined in terms of probabilities $H_p = \{h_p(1), h_p(2), \ldots, h_p(n)\}$ as a function based on the histogram of frequencies $h_p(i) = \frac{h_f(i)}{\sum_{j=1}^{n} h_f(j)}$.

The multi-level thresholding problem consists of obtaining an ordered set $T = \{t_0, t_1, \ldots, t_k, t_{k+1}\}$ of k thresholds, where $1 \leq t_i \leq n-1$ and $t_0 = 0$ and $t_{k+1} = n$ are the boundaries of the histogram, in such a way that some thresholding criterion is optimised. It consists of partitioning a histogram into $k+1$ classes or groups of pixels with consecutive intensity values. In other words, finding an optimal set T of thresholds is equivalent to optimising a function $\Psi'(T) = H_p^k \times [0, 1]^n \to \mathbb{R}^+$ which defines a thresholding criterion over a set of thresholds, where T partitions the set H_p into $k+1$ classes as follows:

$$\begin{aligned}
\zeta_1 &= \{f(x, y)/t_0 = 0 \leq f(x, y) \leq t_1\}, \\
\zeta_2 &= \{f(x, y)/t_1 + 1 \leq f(x, y) \leq t_2\}, \\
&\vdots = \vdots \\
\zeta_{k+1} &= \{f(x, y)/t_k + 1 \leq f(x, y) \leq t_{k+1}\}.
\end{aligned}$$

2.2 Polynomial-time optimal multi-level thresholding

Rueda (2008) proposed a polynomial-time algorithm for multi-level thresholding. This algorithm is polynomial not just on the number of bins of the histogram, but also on the number of thresholds, and runs in polynomial time independently of the thresholding criterion, with Ψ defined as follows:

$$\Psi(T_{0,m}) = \Psi(\{t_0, t_1, \ldots, t_m\}) = \sum_{j=1}^{m} \psi_{t_{j-1}+1, t_j} \tag{1}$$

where $\Psi : H_p^m \times [0, 1]^n \to \mathbb{R}^+ \cup \{0\}, 1 \leq m \leq k+1, t_0$ is the first element of $T, \psi_{l,r} : H_p^2 \times [0, 1]^{l-r+1} \to \mathbb{R}^+ \cup \{0\}$, and for $\psi(l, l+1, \ldots, r, h_p(l), h_p(l+1), \ldots, h_p(r)), \psi(l, r, h_p(l), h_p(l+1), \ldots, h_p(r))$ or for short $\psi_{l,r}$, $l < r$, must satisfy the following three conditions:

- *Condition 1*: For any histogram H_p and any threshold set $T, \Psi > 0$ and $\psi \geq 0$.
- *Condition 2*: For any $m, \Psi(T_{0,m}) = \Psi(\{t_0, t_1, \ldots, t_{m-1}\}) + \psi_{t_{m-1}+1, t_m}$.
- *Condition 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 dynamic programming algorithm can be characterised in terms of solutions to sub-problems. Solving smaller problems in the following order: $\Psi(T_{0,1}), \Psi(T_{0,2}), \Psi(T_{0,3})$, and so on, will avoid resolving any sub-problem two or more times. These smaller sub-problems are incrementally used to solve

larger sub-problems until the whole problem, $\Psi(T_{0,k+1})$, is solved. The dynamic programming algorithm proposed by Rueda (2008) is depicted in Figure 1, where the function $\psi_{l,r}$ corresponds to any thresholding criterion that satisfies conditions 1 to 3. The three main criteria are defined as follows (a complete description of the implementation of these three criteria can be found in the work by Rueda (2008)):

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

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

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

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

It is important to highlight 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 by Rueda (2008) is implemented in our work.

Despite being able to obtain various optimal thresholding sets T, and for various values of k, it is not possible to find the best value of k using the thresholding criterion itself. The following section describes the use of clustering validity indices in order to obtain the best value of k based on the selected thresholding criterion.

2.3 Optimal number of thresholds

The previous thresholding algorithms are not capable of determining the number of thresholds k in which an image can be segmented correctly. However, k has a direct relationship with the number of classes $k+1$, in which a histogram is partitioned by means of a multi-level thresholding algorithm. Viewing thresholding as a problem of clustering pixel intensities, clustering validity indices can be used in order to obtain the best number of classes $k+1$ in which the histogram can be clustered, and hence the number of thresholds. In this work, four clustering validity indices are used to determine the best number of thresholds, which can be found in the work of Maulik and Bandyopadhyay (2002), as follows:

- *Davies-Bouldin Index (DB)*: This index is defined as the ratio between the within-cluster scatter vs. the between-cluster scatter. The goal is to minimise the value of the DB function, which is defined as:

$$DB = \frac{1}{k+1} \sum_{i=1}^{k+1} R_i, \tag{5}$$

where $k+1$ is the number of clusters, $R_i = max_{j \neq i}\{\frac{S_i+S_j}{d_{ij}}\}$ with $1 \leq i \leq k+1$ and $1 \leq j \leq k+1, S_j = \frac{1}{|\zeta_j|} \sum_{i=t_{j-1}+1}^{t_j} h_p(i)\|i-\mu_j\|$ is the within-cluster scatter of cluster ζ_j, and $d_{ij} = \|\mu_i - \mu_j\|$ is the distance between clusters ζ_i and ζ_j.

Figure 1 General algorithm for multilevel thresholding based on dynamic programming

Algorithm *Multilevel_Thresholding*
Input: Probabilities, $H_p = \{h_p(1), h_p(2), ..., h_p(n)\}$. Number of thresholds, k.
Output: A threshold set, $T = \{t_0, t_1, t_2, \ldots, t_k, t_{k+1}\}$.

```
minTj, maxTj ← findThresholdRanges(k)
// Fill columns 1 to k + 1
C(0,0) ← 0; D(0,0) ← 0
for j ← 1 to k + 1 do
  for t_j ← minTj(j) to maxTj(j) do
    C(t_j, j) ← 0;  psi ← ψ_{j,t_j}
    for i ← minTj(j − 1) to min{maxTj(j − 1), t_j − 1} do
      if C(i, j − 1) + psi > C(t_j, j) then
        C(t_j, j) ← C(i, j − 1) + psi
        D(t_j, j) ← i
      end if
      psi ← Compute ψ_{i+2,t_j} from psi, i + 1 and h_p(i + 1)
    end for
  end for
end for
return findThresholds(D)
procedure findThresholdRanges(k: integer)
// Initialize ranges for t_j
for j ← 0 to k + 1 do
  if j = k + 1 then
    minTj(j) ← n
  else
    minTj(j) ← j
  end if
  if j = 0 then
    maxTj(j) ← 0
  else
    maxTj(j) ← n − k + j − 1
  end if
end for
return minTj, maxTj
end procedure
procedure findThresholds(D: table)
T(k + 1) ← n
for j ← k downto 0 do
  T(j) ← D(T(j + 1), j + 1)
end for
return T
end procedure
```

- *Dunn's Index (DN)*: This index is based on the relationship between the cluster sizes and cluster distances. The aim is to maximise the function DN defined as:

$$DN = min_{1\leq i\leq k+1}\left\{min_{1\leq j\leq k+1}\left\{\frac{\delta(\zeta_i,\zeta_{j\neq i})}{max_{1\leq r\leq k+1}(\Delta(\zeta_r))}\right\}\right\} \tag{6}$$

where $\Delta(\zeta_i) = max_{x,y\in\zeta_i}\{d(x,y)\}$ is the diameter of cluster $\zeta_i, \delta(\zeta_i,\zeta_j) = min_{x\in\zeta_i,y\in\zeta_j}\{d(x,y)\}$ is the distance between clusters ζ_i and ζ_j, and $k+1$ is the number of clusters.

- *Index I (IndexI)*: This index is composed of three terms: the number of clusters, the scatter of the clusters and the distance between the clusters. The goal is to maximise the function $IndexI$ defined as:

$$IndexI = \left(\frac{1}{k+1}\frac{E_1}{E_K}D_K\right)^p \tag{7}$$

where $E_K = \sum_{j=1}^{k+1}\sum_{i=t_{j-1}+1}^{t_j} u_{ji}h_p(i)\|i-\mu_j\|$ is the scatter of the clusters, u_{ji} is the membership of intensity i to cluster ζ_j, E_1 is a constant given for the dataset, $D_K = max_{i,j=1}^{k+1}\|\mu_i - \mu_j\|$ is the distance between the farthest clusters, the power p is used to control the contrast between the different cluster configurations (we have taken $p = 2$ in this paper), and $k+1$ is the number of clusters.

- *Calinski Harabasz Index (CH)*: This index is defined as the ratio between the within-cluster scatter and the between-cluster scatter. The aim is to maximise the function CH defined as:

$$CH = \frac{traceB}{k}/\frac{traceW}{n_t - k + 1} \tag{8}$$

where $traceB = \sum_{i=1}^{k+1}|\zeta_i|\|\mu-\mu_i\|^2$ is the sum of between-cluster scatters with μ being the centroid of the entire dataset H_p, $trace\ W = \sum_{j=1}^{k+1}\sum_{i=t_{j-1}+1}^{t_j}\|\zeta_j\|$ $h_p(i)\|i-\mu_j\|^2$ is the sum of within-cluster scatters, n_t is the total number of elements in all clusters, and k is the number of thresholds.

Once the various thresholding criteria and clustering validity indices are defined, we define a way to measure the performance of the combination of both techniques. The following section describes a similarity index which allows us to measure how close are two images segmented in two different ways.

2.4 Manual segmentation vs. automatic segmentation

In order to determine the best combination of thresholding criterion and clustering validity index, different indices can be used. In this section, two of them are discussed, namely the probabilistic Rand index and the sum of squares of relative residuals.

- *Probabilistic Rand Index (RI)*: This index, is simply the percentage of pairs for which there is an agreement. Let $L = \{l_1, l_2, \ldots, l_{NM}\}$ and $L_r = \{l_1^r, l_2^r, \ldots, l_{NM}^r\}$ be the ordered sets of labels l_i for each element of an image defined as a set of ordered pixels without spatial relationship $X = \{x_1, x_2, \ldots, x_{NM}\}$, where NM is the number of pixels in an image, R is the number of reference segmentations L_r to be compared against a specific segmentation L, and $1 \le r \le R$. The RI index is defined as a function over segmented images $RI(L, L_{\{1..R\}}) : L \times L^R \to [0,1]$ which is defined as:

$$RI(L, L_{\{1..R\}}) = \frac{1}{\binom{R}{2}}\sum_{i,j\neq i}^{NM}[I(l_i = l_j)P(l_i = l_j) + I(l_i \neq l_j)P(l_i \neq l_j)] \tag{9}$$

where $l_i, l_j \in L$ are the labels of pixels x_i and x_j, respectively, I is the identity function, $P(l_i = l_j) = \frac{1}{R}\sum_{r=1}^{R} I(l_i^r = l_j^r)$ and $P(l_i \neq l_j) = \frac{1}{R}\sum_{r=1}^{R} I(l_i^r \neq l_j^r)$ are, respectively, the probability that $l_i = l_j$ and the probability that a label $l_i \neq l_j$. This index takes a value one when L and $L_r = \{L_1, \ldots, L_R\}$ are equal, and zero if they do not agree on anything at all.

- *Mean of the Sum of Squares of Relative Residuals (MSSRR)*: This measure is used to evaluate the differences between the threshold levels selected by automatic thresholding and the manual method. The MSSRR is defined in Yang et al. (2001), as:

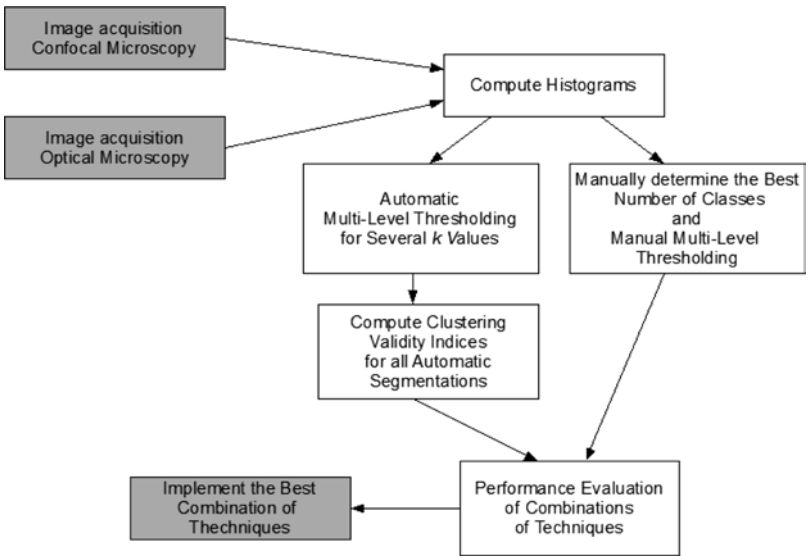
$$\text{MSSRR} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{t_i - t_i'}{t_i}\right)^2 \tag{10}$$

where t_i and t_i' are the thresholds for the ith image found manually and automatically, respectively, and M is the total number of images.

2.5 Determining the best combination of techniques

In order to determine the best combination of techniques for automatic segmentation by means of multi-level thresholding of biofilm images, a method that combines automatic multi-level thresholding and clustering validity indices is proposed. In Figure 2, the general scheme for finding the most appropriate combination of methods is depicted.

Figure 2 General scheme for determining the best combination of techniques for biofilm image segmentation



The first step in the whole process is image acquisition from OM and CLSM and calculating the histogram for each image. Next, a manual multi-level segmentation was performed by means of a trial-and-error process, in order to determine k and

T for each original image. This process is performed by an expert, obtaining the segmented images and the best number of thresholds for each image. Additionally, each original image is automatically segmented by means of the optimal multi-level thresholding algorithm proposed by Rueda (2008) for three thresholding criteria and several values of k (number of thresholds). Clustering validity indices are calculated for each image that is segmented automatically. Finally, the RI index is calculated for each segmented image by means of manual and automatic thresholding, in order to determine the best combination of thresholding criteria and clustering validity indices.

3 Experimental results

A dataset of 649 images was used to evaluate our proposed approach to the segmentation of biofilm images. These images were obtained as follows. Mature biofilms of certain strains of *Pseudomonas syringae* bacteria were developed within the Biofilm and Environmental Microbiology Laboratory (available at http://www.udec.cl/~bem-lab/). The 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. Table 1 shows the features of the images–all of them are 12-bit greyscale images. In addition, the number of thresholds, k, was found manually by a human expert.

Table 1 Dataset of biofilm images used in our experiments

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

Many biofilm images in the dataset are appropriately segmented using a single threshold and a few images are best segmented with more than one threshold. Therefore, the dataset was divided into two subsets of images in order to perform our experiments and to analyse the experimental results. This helps avoid any bias introduced by the difference in the number of images. We should note that for this dataset the images that are best segmented with one threshold are those obtained by CLSM, while those with more than one threshold correspond to the images obtained from OM. However, the relationship between the procedure utilised for acquisition of the images and the appropriate number of thresholds may not necessarily be the case for other images.

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: $\mathrm{RI}^{\mathrm{all}}$ is RI index for all image datasets, $\mathrm{RI}^{\mathrm{clsm}}$ is the RI index for images with one threshold found

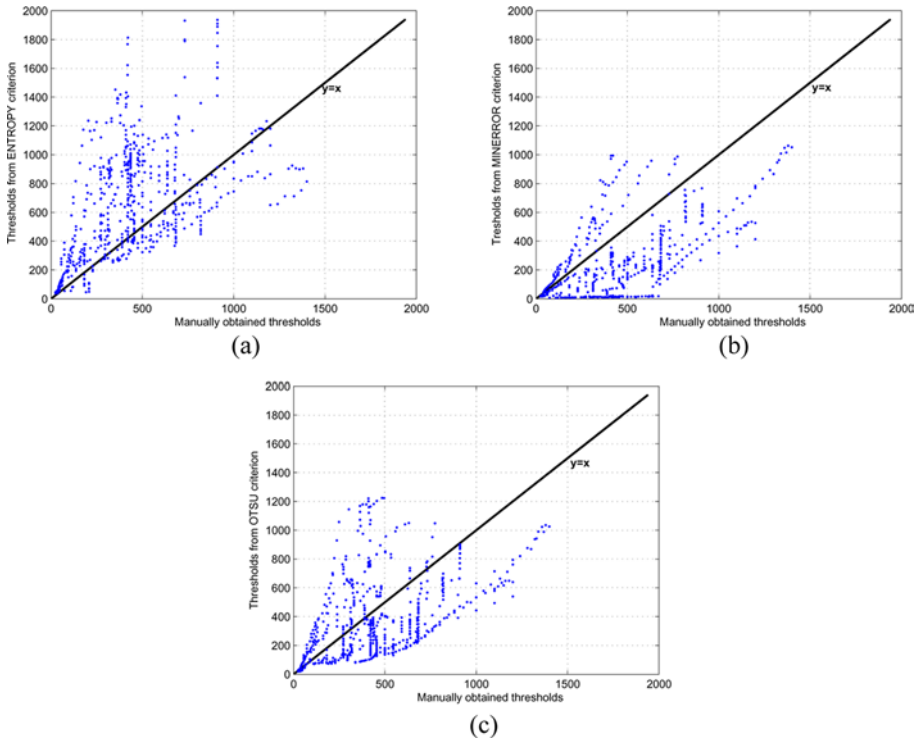
manually, and $\mathbf{RI}^{\mathrm{op}}$ is the RI index for images with more than one threshold found manually. In Table 2, the resulting values for the RI index are depicted for all subsets of images. It is clear that ENTROPY is the best criterion for thresholding images with one threshold. On the 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 datasets of images.

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

Index and dataset	*OTSU*	*ENTROPY*	*MINERROR*
$\mathbf{RI}^{\mathrm{op}}$	**0.7897**	0.7300	0.7713
$\mathbf{RI}^{\mathrm{clsm}}$	0.7283	**0.7767**	0.6086
$\mathbf{RI}^{\mathrm{all}}$	0.6184	**0.7566**	0.5846

Additionally, for images obtained from CLSM microscopy, the MSSRR and the correlation R were calculated in order to evaluate the differences between the threshold levels selected by automatic thresholding and the manual method, where $M = 616$ is the total number of images from CLSM. R was calculated as the Pearson correlation index between the manual and automatic thresholdings. Figure 3(a)–(c) show correlation plots between manually set thresholds and thresholds obtained

Figure 3 Correlation plots between thresholding from automatic criteria and manually set thresholds: (a) correlation for ENTROPY criterion; (b) correlation for MINERROR criterion and (c) correlation for OTSU criterion (see online version for colours)



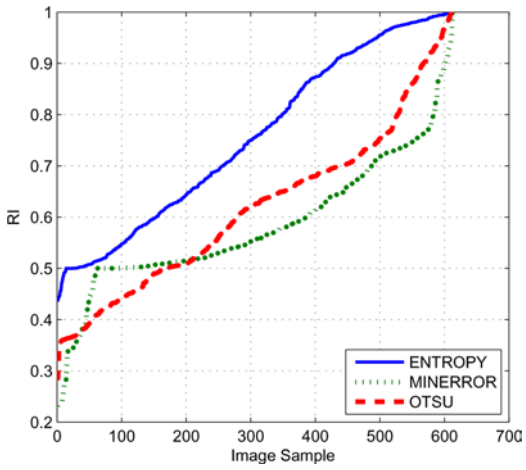
from ENTROPY, MINERROR and OTSU, respectively. For an agreement between automatic and manual thresholding we expect that the points fall on the diagonal line ($y = x$). In Table 3, the resulting values for MSSRR and each thresholding criterion are depicted.

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

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

Figure 4 shows plots of the RI index in increasing order of performance value, based on the number of images being processed (x-axis). It can be observed that the best performance is achieved by the $ENTROPY$ criterion followed by the $OTSU$ criterion.

Figure 4 The RI index for CLSM images, where the number of thresholds was found manually (see online version for colours)



Note the poor performance of the $OTSU$ and $MINERROR$ criteria for segmentation of images with a single threshold. The reason for this is that these criteria are based on clustering algorithms, and hence the presence of one peak only in the histogram make these criteria assign the cluster to the centre of the peak, rather than allocating the threshold on *one side* of that peak.

As opposed to this, the $ENTROPY$ criterion attains better results for one-peaked histograms. This is reasonable, since the formulation of this criterion leads to assigning lower entropy values when a class distribution is *flat*, while when the peaks are higher, the entropy is higher too. This implies that the $ENTROPY$ criterion leads to similar results to those obtained by manual thresholding.

3.2 Finding the best number of thresholds

We have considered a special measurement on the error in estimating the number of thresholds for the complete dataset of images. Table 4 shows the Mean Square Error

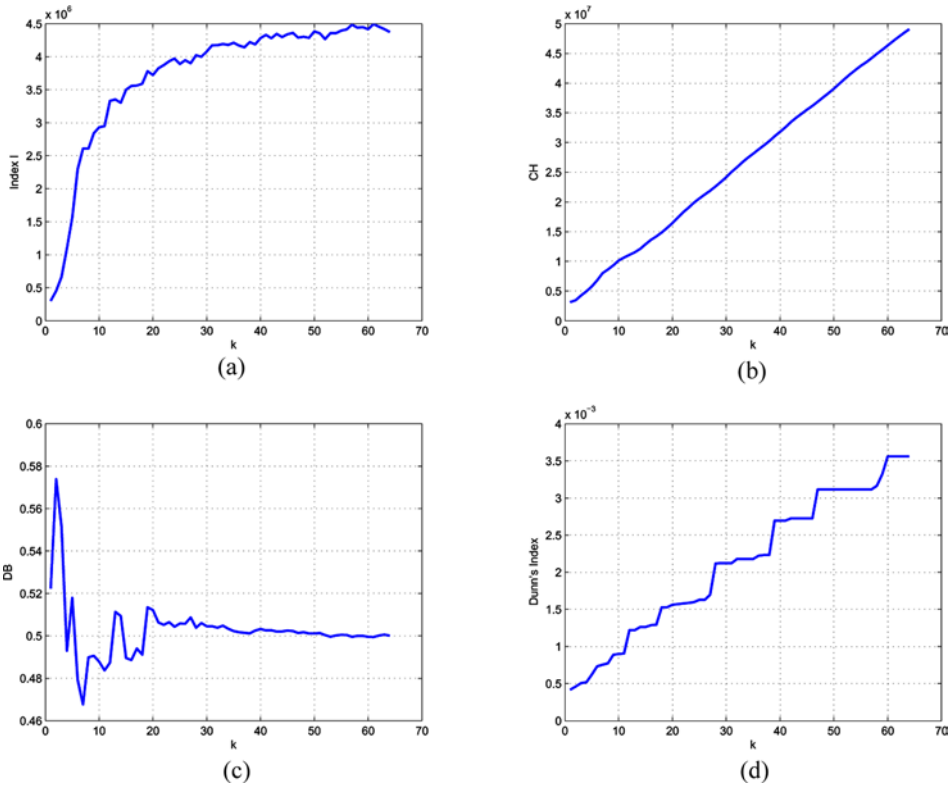
(MSE) for each combination of thresholding criteria and clustering validity indices. The DB index achieves the best performance with the $ENTROPY$ criterion, which, again, indicates that the combination $ENTROPY + DB$ attains a very good performance in most of the cases for different datasets of images.

Table 4 Mean Squared Error (MSE) for the estimation of the best number of thresholds

	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

The clustering validity indices have a direct relationship between them in their formulation; however, each index has a different behaviour depending on the number of thresholds selected. The behaviour of each validity index can be observed in Figure 5. Although the plots are for one of the images in the dataset, they represent the behaviour of the clustering validity indices for the entire dataset.

Figure 5 General behaviour of clustering validity indices: (a) IndexI; (b) CH; (c) DB and (d) DN (see online version for colours)



In Figure 5, one observes that the indices *IndexI, CH* and *DN* are (for most of the values of k) monotonically increasing functions of k (Figure 5(a), (b) and (d)), yielding the best performance when the function achieves the maximum, i.e., when $k = 61, k = 64$ and $k = 64$, respectively. This behaviour does not provide a clear direction on how to determine what is the optimal number of clusters with which the image should be segmented, and illustrate 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. Thus, 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 produce 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 (Figure 5(c)).

In addition, $IndexI$ did not lead to a good performance because the term E_K significantly reduces its value when the number of clusters grows. This situation occurs because the dispersion of the data from the centres of the classes is smaller when the values of k are larger, and the value of the distance between each point and the centre becomes increasingly small since the classes have fewer members and these members are closer when dealing with one dimension.

On the other hand, the DN index grows when k grows. This is because the diameter of the clusters is smaller when k is larger. In addition, the distance between the clusters is defined as the shortest distance between two members of each cluster. Then in the case of a one-dimensional histogram, this corresponds to clusters of intensities located next to each other, and hence this value has no impact on the value of the index, and the index is almost entirely dominated by the diameter of each cluster, which decreases when k increases.

The DB index is defined as the sum of the dispersion of pairs of clusters divided by the distance between them, implying that the index is lower when the dispersion of pairs of clusters is smaller compared to their distance, which does not have a direct influence on the number of clusters. This index has a behaviour mainly dominated by the nature of the clusters, favouring distant and compact clusters, which is equally desirable in the separation of the *peaks* in a one-dimensional histogram.

3.3 Performance for segmentation with cluster validity indices

Table 5 shows the values of the RI index for all biofilm images. From this test, the values included in Table 5 show that the best combination is ENTROPY + DB for RI^{all} index. Also, it is clear that the thresholding criterion 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

3.4 One threshold

All biofilm images obtained by the confocal microscope have a single optimal threshold (manually found by the expert). Table 6 shows the performance of different thresholding criteria and clustering validity indices for the image segmentation of biofilms 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 thresholding 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 does not differ significantly. However, it is clear that the performances of thresholding criteria are significantly influenced by the number of clusters 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

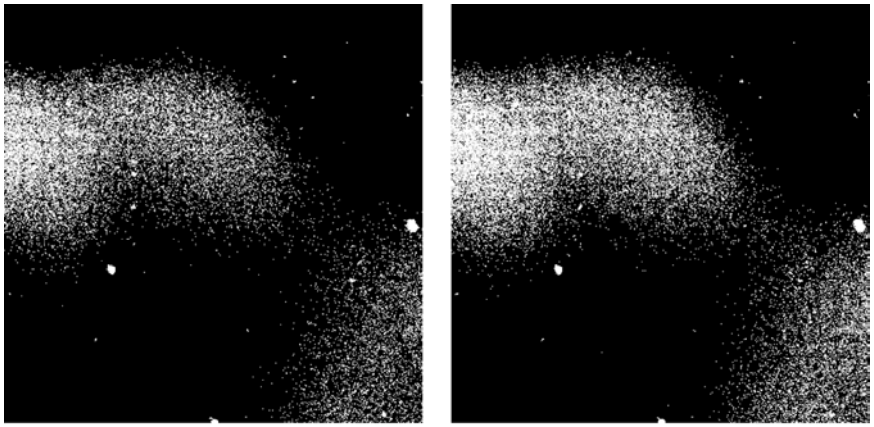
3.6 Visual validation

3.6.1 Image segmentation over isolated images

Figure 6(a) shows the manual segmentation of a biofilm image with one threshold compared with the automatic segmentation using the $ENTROPY$ criterion and the DB clustering validity index (Figure 6(b)). The automatic segmentation sets the

thresholds with a value slightly lower than the manual segmentation, implying that more pixels with high intensities are labelled with white colour. As seen in the figure, automatic segmentation produces results quite similar to those of the segmentation performed by the expert, and even providing more level of detail.

Figure 6 Binary segmentation of a CLSM image: (a) segmented manually and (b) segmented automatically



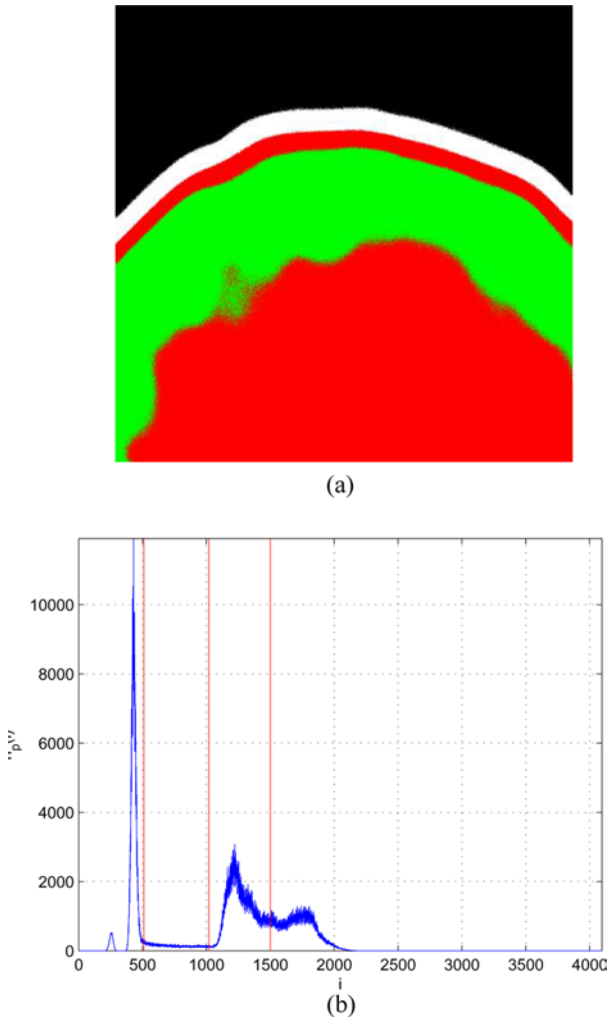
Figures 7(a) and 8(a) show the manual segmentation of a biofilm with more than one threshold compared to the automatic segmentation that combines $ENTROPY + DB$ respectively. As can be seen in Figure 8(b), the result of automatic segmentation is close to that of manual segmentation (Figure 7(b)), setting the thresholds to almost the same values when segmentation is done by the expert.

3.6.2 3D reconstruction of the biofilm

Rebuilding the structure of a biofilm from image stacks of confocal microscopy offers a powerful visualisation tool which allows observing the images from different angles. Figures 9(a) and 10(a) 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 $ENTROPY + DB$. As can be seen, the image reconstructed automatically is quite similar to the manual rebuilding done by an expert.

The existence of cross-sectional images in the biofilm also makes it possible to use reconstruction techniques for 3D images. One of the most common reconstruction algorithms are the ones based on *marching cubes*, proposed by Lorensen and Cline (1987). This algorithm requires as input a cloud of points (ideally binary). Then, the algorithm connects these points between adjacent surfaces incrementally by means of a pre-determined set of rules based on the type of distribution of the points in a cube. Figure 11 shows the 3D reconstruction of a biofilm over 50 cross-sectional images. As can be seen in the figure, the properties of luminosity of scene and perspective of vectorial images enhances substantially the visualisation of the structure of the biofilm.

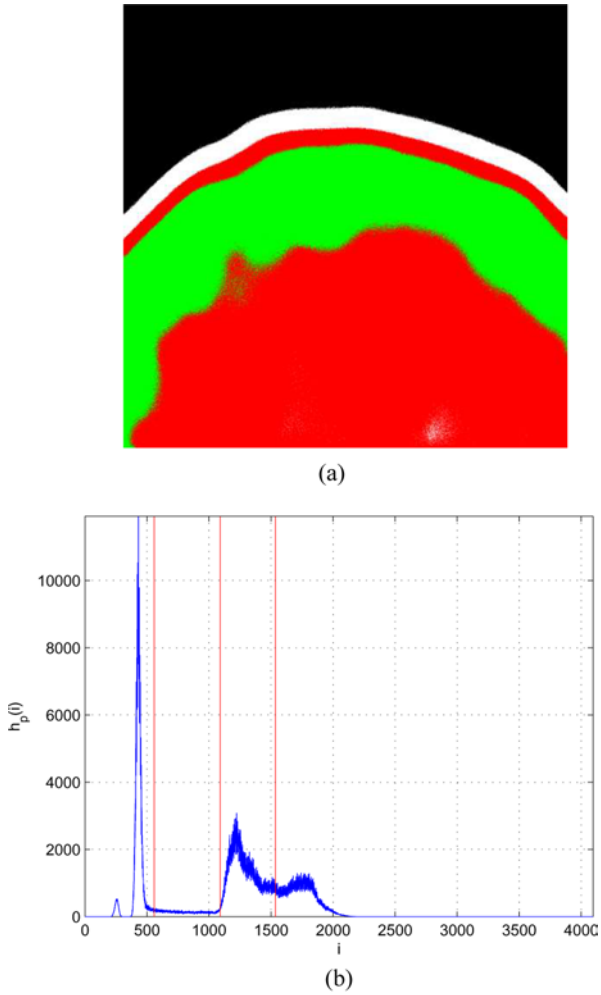
Figure 7 Multi-level thresholding segmentation: (a) optical image segmented manually and (b) histogram of (a) (see online version for colours)



4 Quantification validation

A quantification process was also done in a laboratory. In Table 8, we show the ratios of live/dead cells for two biofilms with three days of growth (3D biofilms) and four days of growth (4D biofilms). Ratios in the last row are obtained from laboratory measurements. We applied the biofilm quantification techniques, described in the work of Beyenal et al. (2004b), on stacks of CLSM images (of these biofilms) automatically segmented by means of ENTROPY, MINERROR and OTSU criteria, where the best number of thresholds was found by using the DB index. Thus, using the methods proposed by Beyenal et al. (2004b), we were able to estimate the ratios of life/dead cells of biofilm images after segmentation and three-dimensional reconstruction, and then compare them with those obtained in the

Figure 8 Multi-level thresholding segmentation: (a) optical image segmented automatically and (b) histogram of (a) (see online version for colours)



laboratory. As can be seen in the table, the ENTROPY criterion gave results closest to laboratory values. MINERROR and OTSU clearly underestimate the laboratory ratios by a large margin, confirming again their poor performance in the RI results in Table 5.

5 Conclusions

A new technique for automatic segmentation of biofilm images has been proposed, which is based on efficient and optimal multi-level thresholding. The algorithm runs in polynomial time complexity and allows to estimate the number of thresholds through the well-known Davies Bouldin (DB) clustering validity index. The thresholding criterion that performs as good as the manual segmentation was found

Figure 9 Biofilm reconstruction: (a) manual reconstruction and (b) automatic reconstruction (see online version for colours)

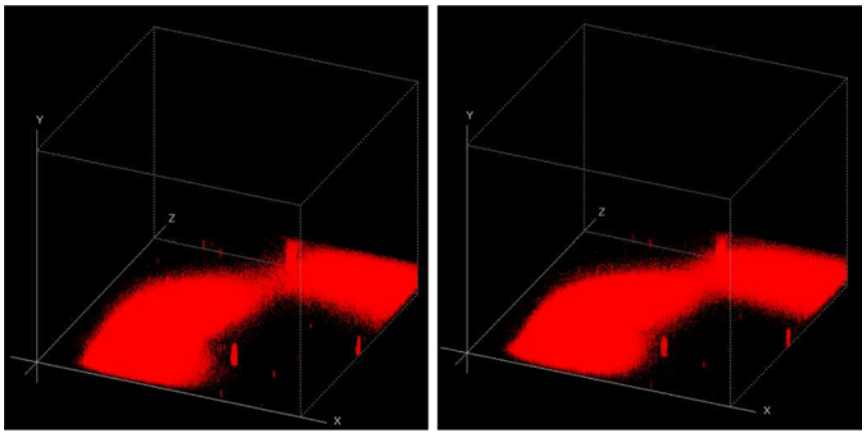
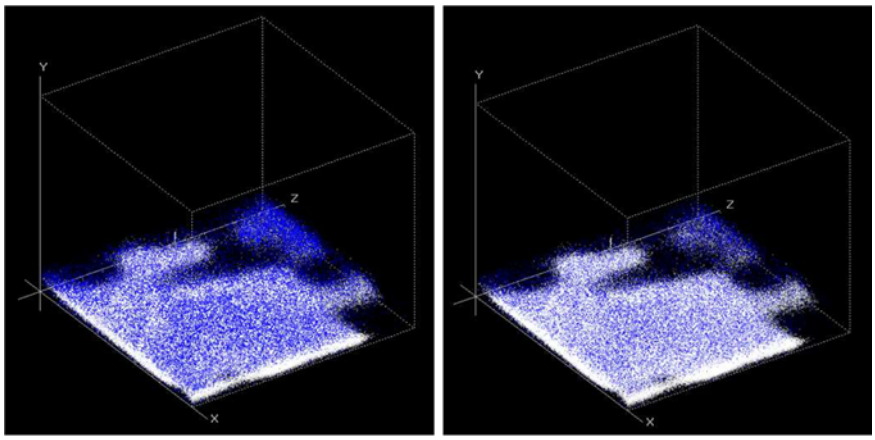


Figure 10 Biofilm reconstruction: (a) manual reconstruction and (b) automatic reconstruction (see online version for colours)



to be *maximum entropy*. This was assessed by using the objective measure of *Probabilistic Rand Index*, which compares segmentations done by the proposed method against the segmentation done by a human expert. To validate our method, we compared the ratio counts of live and dead cells in biofilms between laboratory measurements and image quantification (using our methods); we obtained the best results with the combination $ENTROPY + DB$.

Future work involves quantification of biofilms that enables researchers in the field to create an objective judgement about the relevant features of a biofilm and between different biofilms, by measuring the coverage area, porous surface, spatial correlation, entropy, among others. On the other hand, vectorisation of segmented images could facilitate visualisation, allowing to *navigate* through the biofilm structure in three dimensions and helping observe the internal structure of the biofilm. These are issues that remain to be investigated.

Figure 11 Biofilm reconstruction by marching cubes

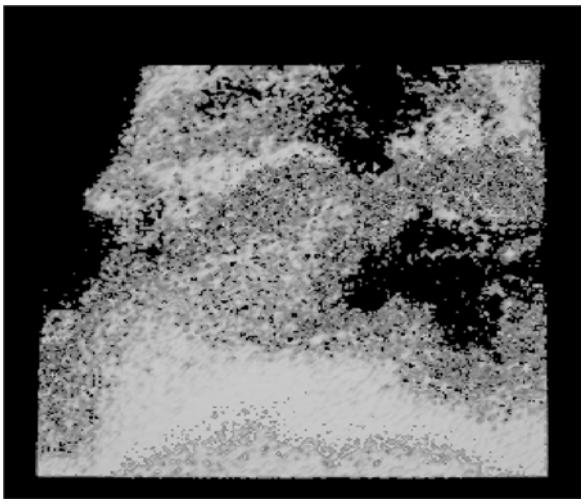


Table 8 Quantification comparation between cells recounting and biovolume quantification

	3D Biofilm	*4D Biofilm*
ENTROPY	**89.3/10.7**	**92.4/7.6**
MINERROR	46.1/53.9	50.2/49.9
OTSU	56.4/43.6	61.6/38.4
Laboratory	**91.3/8.7**	**98.6/1.4**

Acknowledgements

This work has been partially supported by NSERC, the Natural Sciences and Engineering Research Council of Canada (Grants No. RGPIN261360 and RGPIN228117), the Canadian Foundation for Innovation (Grant No. 9263), the Ontario Innovation Trust, and the University of Atacama (University Grant for Research and Artistic Creativity, Grant No. 221172).

References

Beech, I.B., Sunner J.A. and Hiraoka, K. (2005) 'Microbe-surface interactions in biofouling and biocorrosion processes', *International Microbiology*, Vol. 8, No. 3, pp.157–168.

Beyenal, H., Donovan, C., Lewandowski, Z. and Harkin, G. (2004a) 'Three-dimensional biofilm structure quantification', *Journal of Microbiological Methods*, Vol. 59, No. 3, pp.395–413.

Beyenal, H., Lewandowski, Z. and Harkin, G. (2004b) 'Quantifying biofilm structure: facts and fiction', *Biofouling*, Vol. 20, No. 1, pp.1–23.

Beyenal, H., Tanyolaç, A. and Lewandowski, Z. (1998) 'Measurement of local effective diffusivity in heterogeneous biofilms', *Biotechnology and Bioengineering*, Vol. 56, Nos. 8–9, pp.656–670.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R. and Lappin-Scott, H.M. (1995) 'Microbial biofilms', *Annual Review of Microbiology*, Vol. 49, No. 1, pp.711–745.

Hall-Stoodley, L., Costerton, J. and Stoodley, P. (2004) 'Bacterial biofilms: from the natural environment to infectious diseases', *Nature Reviews Microbiology*, Vol. 2, No. 2, pp.95–108.

Heydorn, A., Nielsen, A.T., Hentzer, M., Sternberg, C., Givskov, M., Ersboll, B.K. and Molin, S. (2000) 'Quantification of biofilm structures by the novel computer program comstat', *Microbiology*, Vol. 146, No. 10, pp.2395–2407.

Johnson, L.R. (2008) 'Microcolony and biofilm formation as a survival strategy for bacteria', *Journal of Theorical Biology*, Vol. 251, No. 1, pp.24–34.

Jorgensen, T.M., Haagensen, J., Sternberg, C. and Molin, S. (2003) 'Quantification of biofilm structure from confocal imaging', *Visionday 2003 of Technical University of Dinmark*, Lyngby, Denmark.

Klapper, I. (2006) 'Effect of heterogeneous structure in mechanically unstressed biofilms on overall growth', *Bulletin of Mathematical Biology*, Vol. 66, No. 4, pp.809–824.

Lewandowski, Z., Altobelli, S., Majors, P. and Fukushima, E. (1992) 'NMR imaging of hydrodynamics near microbially colonized surfaces', *Water Science and Technology*, Vol. 26, pp.577–584.

Lorensen, W.E. and Cline, H.E. (1987) 'Marching cubes: a high resolution 3d surface construction algorithm', *ACM SIGGRAPH Computer Graphics*, Vol. 21, No. 4, pp.163–169.

Maulik, U. and Bandyopadhyay, S. (2002) 'Performance evaluation of some clustering algorithms and validity indices', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 12, pp.1650–1655.

Merod, R.T., Warren, J.E., McCaslin, H. and Wuertz, S. (2007) 'Toward automated analysis of biofilm architecture: bias caused by extraneous confocal laser scanning microscopy images', *Applied and Environmental Microbiology*, Vol. 73, pp.4922–4930.

Nikolaev, Y.A. and Plakunov, V.K. (2007) 'Biofilm–city of microbes or an analogue of multicellular organisms?', *Microbiology*, Vol. 76, No. 2, pp.125–138.

Otsu, N. (1979) 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, No. 1, pp.62–66.

Rueda, L. (2008) 'An efficient algorithm for optimal multilevel thresholding of irregularly sampled histograms', *Proceedings of 7th International Workshop on Statistical Pattern Recognition*, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, Orlando, Florida, USA, pp.612–621.

Stewart, P.S. and Franklin, M.J. (2008) 'Physiological heterogeneity in biofilms', *Nature Reviews Microbiology*, Vol. 6, No. 3, pp.199–210.

Tolle, C.R., McJunkin, T.R. and Stoner, D.L. (2003) *Mapper: A Software Program for Quantitative Biofilm Characterization*, Technical Report, Idaho National Engineering and Environmental Laboratory, Idaho Falls, USA, ID 83415-2210.

Yang, S. and Lewandowski, Z. (1995) 'Measurement of local mass transfer coefficient in biofilms', *Biotechnology and bioengineering*, Vol. 48, No. 6, pp.737–744.

Yang, X., Beyenal, H., Harkin, G. and Lewandowski, Z. (2000) 'Quantifying biofilm structure using image analysis', *Journal of Microbiological Methods*, Vol. 39, No. 2, pp.109–119.

Yang, X., Beyenal, H., Harkin, G. and Lewandowsi, Z. (2001) 'Evaluation of biofilm image thresholding methods', *Water Science and Technology*, Vol. 35, No. 5, pp.1149–1158.