

Dynamic Feature Fusion in the Self Organising Tree Map - Applied to the Segmentation of Biofilm Images

Matthew Kyan
School of Electrical and Information
Systems Engineering
University of Sydney
NSW 2006 Australia
E-mail: mkyan@ee.ryerson.ca

Ling Guan
Department of Electrical and
Computer Engineering
Ryerson University
Toronto ON M5B2K3 Canada
E-mail: lguan@ee.ryerson.ca

Steven Liss
Department of Chemistry
and Biology
Ryerson University
Toronto ON M5B2K3 Canada
E-mail: sliss@ryerson.ca

Abstract—The Self Organising Tree Map (SOTM) neural network is investigated as a means of segmenting micro-organisms from confocal microscope image data. Features describing pixel & regional intensities, phase congruency and spatial proximity are explored in terms of their impact on the segmentation of bacteria and other micro-organisms. The significance of individual features is investigated, and it is proposed that, within the context of micro-biological image segmentation, better object delineation can be achieved if certain features dominate the initial stages of learning. In this way, other features are allowed to become more/less significant as learning progresses: as the network gains more knowledge about the data being segmented. The efficiency and flexibility of the SOTM in adapting to, and preserving the topology of input space, makes it an appropriate candidate for implementing this idea. Preliminary experiments are presented and it is found that favouring intensity characteristics in the early phases of learning, whilst relaxing proximity constraints in later phases of learning, offers a general mechanism through which we can improve the segmentation of microbial constituents.

I. INTRODUCTION

Microscopy has long been considered an essential investigative tool in the study of complex micro-biological environments, such as those encountered in modern Biofilm research. The term *Biofilm* is used to describe a collection of micro-organisms anchored to some form of substrata [1], which may exist in a variety of different situations – from the dental plaque found on teeth, to microbial *flocs* (highly complex aggregations of microbial cells or bacteria, bioorganic and inorganic material) suspended in aqueous environments [2]. In such environments, physical structure plays a crucial role in the many key processes that occur.

Knowledge of the structural factors affecting such processes becomes important in formulating solutions to many pertinent environmental problems. In the

performance of waste water treatment, for instance, a process known as *activated sludge* is utilised to promote the stabilised formation of flocs. This process then aids in the separation of suspended microbial contaminants from treated effluent. A common problem occurs in such systems when flocs form with poor settling properties, thereby degrading the overall treatment process [2]. To gain insight into the structural properties and microbial behaviour in this and many other environmental applications, microscopy remains the major tool through which non-invasive study may be conducted *in-situ* at the cellular level.

Major technological advances in microscopy such as the introduction and development of the confocal laser scanning microscope (CLSM), can now overcome many of the normal problems associated with using standard instrumentation to capture images of these, typically dense, heterogeneous environments. The ability for CLSM to probe into the depth of a specimen and record a stack of 2D spatially registered image slices allows for reconstruction of the 3D structural information microbiologists and engineers need to form better models of the biofilms or flocs under investigation. At this point, biofilm structure is understood more in a qualitative rather than a quantitative sense [3]. As such, the linking of parameters characterising biofilm performance, with structural features remains a hot topic of research.

In this work, we identify that much of the current approaches to quantifying biofilm structure hinge on consideration of global characteristics. Such measures break down when the image data is extremely heterogeneous due to information lost through crude segmentation that is not representative of the internal components of the data – a factor which is particularly evident in images of microbial floc. We thus present a neural network approach to the segmentation of the internal constituents of biofilm/floc images – incorporating and fusing a set of image features that together, cover a

range of key aspects of the human visual system (HVS). Finally, we introduce a refinement based on the dynamic adjustment of input feature significance to guide the SOTM as it searches through feature space for a better perceptual representation of the input data.

II. THEORETICAL CONSIDERATIONS

A. Limitations of Current Approaches

At present, the majority of available techniques attempt to quantify biofilm structure based on an initial, crude thresholding (often manual). Many automatic techniques have been proposed, but lack robustness, especially when applied to images of extremely heterogeneous biofilm: a situation magnified in the case of floc studies. Popular approaches then generally measure morphological and textural parameters [3]. Morphological measures discussed in the literature are dependant almost exclusively on an initial thresholding. Textural approaches tend to apply a global quantification measure over the entire image, effectively averaging over any true heterogeneity existing in the data.

As such, better forms of segmentation are required to, firstly, achieve more robust and hence repeatable isolation of biofilm/floc structure from a background of noise imparted through the imaging optics; secondly, to isolate the internal constituents of floc, such that parameters measured may more appropriately characterise the distribution and nature of materials present.

In this regard, it should be acknowledged that quite often, epi-fluorescent modes of the microscope are used to image the distribution of molecular markers introduced to a specimen. In such applications, the markers bond with particular molecules present in structures of interest so that they can be isolated from their surroundings. However, often these molecules exist in many different materials present, and generally are present in varying concentrations, reflected as a range of graylevels in the image. In such instances, segmentation via crude thresholding of foreground from background will result in the loss of much, potentially significant information.

B. Segmentation with the SOTM neural network

In this current work, a tree structured neural network known as the Self Organising Tree Map (SOTM) is investigated and developed toward the goal of achieving better segmentation of biofilm constituents. The SOTM partitions a feature space description of input data, by locating clusters (represented by neurons) of high density within this feature space. Like its counterpart, the self organising map (SOM), competitive learning is used to locate clusters such that the final representation maintains the topology of the feature space, yet does so in a more flexible and efficient manner by dynamically generating a

model of this topology as it parses the input space. This results in a representation that tends not to suffer (as SOM does), from neurons being trapped in regions of low density [4,5].

1) *SOTM Architecture*: The process for SOTM learning is outlined in the following algorithm:

- Initialise the root neuron with a randomly selected training vector from feature space X .
- Randomly select a new feature vector x , and compute the Euclidean distance, d_j , to all currently existing neurons w_j ($j=1, \dots, J$), where J is the total number of neurons currently allocated
- Select the winning neuron w_k , with minimum d_j : $d_j(x, w_k) = \min_j d_j(x, w_j)$
- If $d_j(x, w_k) \leq H(t)$, then update the weight vector of the k^{th} neuron using the reinforced learning rule: $w_k(t+1) = w_k(t) + \alpha(t)[x - w_k]$; where $\alpha(t)$ is the learning rate and $H(t)$ is a hierarchy control function. Both functions decay with time: $\alpha(t)$ lies on $[0,1]$, whilst $H(t)$ controls the levels of the tree.
- Alternatively, if $d_j(x, w_k) > H(t)$, spawn a new neuron from w_k at the position x and reset $\alpha(t)$.
- Continue from step b) until all neurons are allocated and/or there is no significant change in the SOTM.

As indicated above, the network is initialised with the random selection of a vector from input space (stimulus): typically each pixel will be represented by a single input vector. The vector may be as simple as a single graylevel intensity value, or it may be a more complex multidimensional vector with each dimension representing a different pixel feature (such as some form of local statistic or otherwise).

Learning progresses through the random selection of new stimuli from the feature space. With each, the currently existing neurons in the network compete to see which is the closest to the stimuli in a Euclidean sense. We can think of this as stimulation of the most similar representation in the memory of the network.

Once a neuron is stimulated, a decision is made as to whether the currently chosen neuron is significantly different from the input or not. This typically occurs in the earlier stages of network growth when there are fewer neurons that exist to compete. The hierarchy control function (described in section II.B.2) is responsible for making this decision. If a neuron is found to exist that is significantly similar to the stimulus, it 'learns' some information from that stimulus. This is achieved by adjusting the neuron (memory) vector toward the current stimulus by a certain proportion $\alpha(t)$ of the Euclidean distance between them.

If there is no significant similarity, then the network figures that it needs to allocate a new neuron (cluster) to the network topology. This neuron then becomes a child of

the neuron it was found to be closest to. Over time, as new neurons are allocated to represent the input, more competition for each new stimulus occurs.

The learning rate $\alpha(t)$ is allowed to decay over time as more neurons are allocated – thus ensuring convergence in that the neurons will eventually settle and ‘tune’ themselves to represent the centres of dense clusters occurring in the feature space. The hierarchical function also decays, allowing for neurons to be allocated as leaf nodes of their closest neurons from the previous generation (previous state of the network). Thus the SOTM forms a flexible tree structure that spreads and twists across the feature space, converging to what is essentially, a topologically aware k-means representation.

2) SOTM Partitioning of Feature Space: As mentioned in section II.B.1, the generation of new neurons is controlled in essence an exponentially decaying hierarchy control function $H(t)$. It acts more or less as a radius (or more correctly, an ellipsoid) of significance over the feature space: emanating from the neuron found to be closest to the current stimulus. This is depicted in the two and three dimensional feature space examples of Fig. 2.

If the current stimulus is within the range defined by this radius (within the ellipsoid boundary), then it is considered significantly similar, and that neuron learns some information from the stimulus. If the stimulus occurs outside the boundary of significant similarity, then the network will spawn a new neuron, attaching it to its closest ‘parent’ neuron.

The partitioning that ensues can be visualised in Fig. 2(a). With the root neuron allocated, stimuli falling within a large radius of significance cause the root neuron to perform a random walk about the centre of density of the entire feature space (i.e. its expected value).

When a random stimulus eventually falls outside the radius, a new neuron is spawned and the feature space is instantly partitioned into regions A and B. With this partitioning, the two neurons now compete, and reconverge to their new regional centres of density.

The timing diagram in Fig. 3 shows an example of how the hierarchical function decays, and subsequent neurons are generated. Current work has revealed that a stepped version of the decaying $H(t)$ function is more desirable. This allows for an adequate search of the input space at a fixed radius, before considering a new range of candidates for neuron generation. Existing neurons are thus given adequate time to readjust to the latest partitioning.

It is noted at this point, for reference, that in section II.B.4, we shall suggest what should happen when we bias the Euclidean distance calculation considered in the competitive learning phase, in any one feature dimension. At this point, we might expect that if one particular dimension is weighted down with respect to others, then stimuli distributed along that dimension would be considered to be much closer to the closest neuron than

they actually are. As a result, partitioning along other dimensions would occur much more readily.

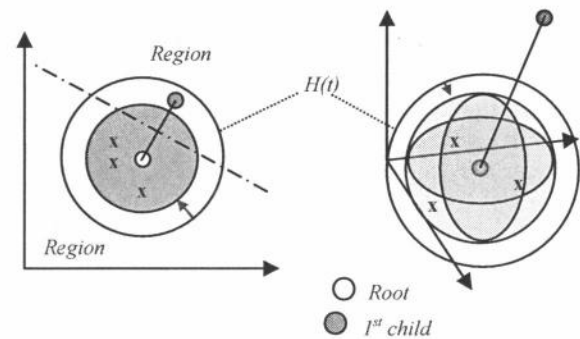


Fig. 2. (a) 2D (left) and (b) 3D (right) feature spaces. x's represent initial stimuli falling within the radius/ellipsoid of significant similarity

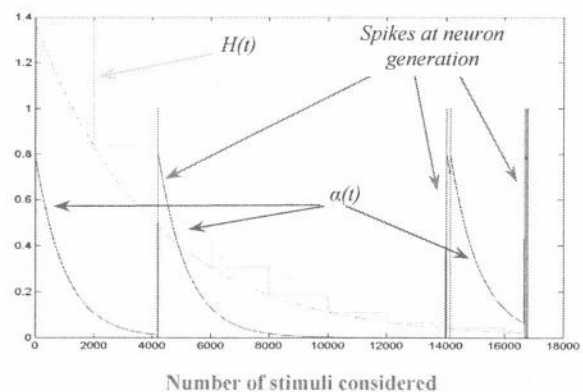


Fig. 3. Timing waveforms governing the evolution of the SOTM

3) Feature Selection and Fusion: In exploring the effectiveness of the SOTM for segmentation, a set of features is fed into the network, each describing a different aspect of the input data. In this work we have chosen a small set of what the authors believe are features fundamental to a basic description of objects in a graylevel scene, as segmented by the HVS. These features include: Graylevel (GL) of an individual pixel, which describes highly localised intensity variation within an object; Average graylevel of a disk shaped region (GL_{disk}) surrounding the current pixel, which describes more regional variation across an object; Phase Congruency (PC) [6], which is more akin to describing the edge information present in an image, however, unlike traditional edge detectors, PC has the added advantage of incorporating scale information into its calculation, thus tends to highlight significant features that the HVS is thought to respond to; Position (XY), which simply relate to the positional co-ordinates (x,y) of pixels in the image plane, considered here such that partitioning across position will give formed clusters a sense of proximity within the image.

Of course, in reality it is highly likely that there are many additional features and factors that contribute to object discrimination as performed by the HVS.

We also note that, in this application, although we primarily need to segment 3D data obtained from the confocal microscope, we restrict our experiments for the current work to the 2D segmentation of an individual slice, so that the significance of these features with respect to one another (as fused through the SOTM), can be assessed, along with the impact of proposed refinements to the SOTM model.

4) *SOTM Model Refinement through Dynamic Weighting of Feature Significance:* As a final step, based on the results of applying fixed global weightings to each of the above features for SOTM clustering, we suggest that the network's ability to cluster across an input data space – whilst efficient and optimal in a vector quantisation sense – may not necessarily converge to locate clusters that represent objects as perceived by the HVS. We propose that in the natural course of learning, certain features may become more (or less) significant. This would have the effect of a non-linear biasing or warping of the final topology of the network away from the vector quantised topology. The resulting topology potentially offering a more object/region based partitioning of the input feature space, and thus, more of an object based segmentation suitable for the delineation of constituent materials existent in biofilm/floc images.

Theoretically, there are quite a number of parameters one could vary in the competitive search, learning, hierarchical function decay, or clustering phases of SOTM segmentation. One might even apply weightings to selectively control all such parameters differently. In this work however, we introduce a dynamic set of weights in the competitive search phase of learning only. We start by applying a set of initial feature weights (β), to the competitive Euclidean distance measure in the SOTM algorithm (section II.B.1) such that step (c) now becomes:

$$d_j: d_j(\mathbf{x}, \mathbf{w}_i) = \min_j [\beta \cdot \gamma(t) \cdot (d_j(\mathbf{x}, \mathbf{w}_i))] \quad (1)$$

where $\gamma(t)$ denotes a set of exponential functions that force individual feature weights to grow, decay, or remain constant throughout learning.

It follows that if individual weights are dynamically increased or decreased gradually (toward the weights of the other features) over the entire learning process, then partitioning would begin by favouring the highest weighted feature(s), shifting toward the other features as they, in turn, become (relatively) more significant.

If feature weights cross over in value, the ratio between them would also switch, thus promoting the shift of partitioning from one feature to another. During this progression, neurons are being allocated in a manner that would favour the highest weighted feature dimension at any one time during learning. This might be a significant dynamic in terms of object segmentation, where neurons to

be allocated could favour one dimension until sufficient information is learnt from that dimension, before shifting focus to another dimension.

Such concepts are explored in the following experiments to draw conclusions as to how such dynamics might result in more effective object segmentation as applied to complex, heterogeneous microbiological image data.

III. RESULTS AND DISCUSSION

Experiments were conducted on individual 2D confocal image slices from a sample of FITC-stained biofilm. Of the slices, one was chosen for this presentation, such that adequate comparison could be drawn as to the impact of feature adjustment on the resulting SOTM segmentation. A number of experiments were conducted, and in each, constraints were imposed both in the sets of features to be fused by the SOTM, and by imposing a bias or weighting to the competitive search phase of SOTM learning (step (c) in the SOTM algorithm). Three main dynamics were explored.

The first involved the dynamic of feature space definition. In this test, our goal was to consider how the SOTM responded when features were added to the feature space description, one by one. With each new definition, we evaluate the impact on segmentation. A 64 neuron SOTM was utilised in this process.

Feature spaces (FS_i), were defined as follows, with reference to section II.B.3, whilst the segmentation results are presented in Fig. 4 (a) $FS_1=(GL)$; (b) $FS_2=(GL, GL_{dsk})$; (c) $FS_3=(GL, GL_{dsk}, PC)$; (d) $FS_4=(GL, GL_{dsk}, PC, X, Y)$. We consider both X and Y to be of equal significance, and to be used in unison to represent position of a pixel. All features remained unbiased throughout learning, and were normalised to eliminate any possibility of natural ranges in the data adversely affecting the outcome of the segmentation. Two images are shown for each case, left images represent the segmented dataset labelled with the GL component of the classifying neuron. Since the analysis is subjective, we offset this image with a randomly colour-labelled version, such that any assessment of region/object localisation is not biased by our own tendency to segment the resulting GL images naturally using our own built in HVS.

As can be seen from the labelled images on the right, both (a) and (b) demonstrate little in the way of a contiguous region (connected set of labels), with the exception of regions of extreme uniformity in GL existing in the background. With PC, in (c), we start to observe some mild regional associations. The overall segmentation however is still rather contaminated by the signal based pixel statistics, each competing (in an equal fashion) for the right to represent the input space. If one considers the PC feature, in being more representative of significant features and edge information, one might assume that a large dynamic range of PC values might exist, but over a

more limited expanse of the input data set. As such the network would tend to encounter large differences in PC more often, thus partition in this dimension more readily. Even with this though, the majority of the labelled segmentation appears quite random and with little grouping across regions.

When adding position to the mix, the SOTM appears to immediately discover more regional based associations. Whilst not delineating all structures per se (limited by its restriction to 64 neurons maximum), a good number of neurons appear to favour a relatively uniform partitioning of XY space. One would assume that the same occurs in the other dimensions (based on the relative density of information in those dimensions). In the XY plane, the density is always uniform, as the X & Y vectors carry no information on their own. The partitioning in this plane however, allows for neurons to separate and yet still learn from GL regional information. This could offer an insight into how separate notions of objects may be formed within the memory of the SOTM, due to their separation in proximity. At any rate, it seems clear that there is strong interplay between the knowledge of signal characteristics and proximity, and the SOTM demonstrates potential in differentiating this.

In the second experiment, we consider applying a bias to the signal & spatial components of feature space. In separate runs of a 16 node SOTM, we adjust the significance weighting of XY components, to investigate how the segmentation is impacted when the competitive learning process is biased toward different features. The results are presented in Fig. 5 (a)-(c). In all cases: GL, GL_{dsk} & PC remain unbiased (weighting of 1), whilst XY is changed. In (a), the weighting for XY is 0.01, in (b) 0.02, and (c) 0.1. A high ratio of Signal vs. XY weightings favour a relatively weaker partitioning in the XY plane. Thus the number of neurons split across (GL, GL_{dsk}, PC) will tend to outnumber those split across XY. Sensitivity of the SOTM to proximity is then considered via a range of fixed XY weightings.

From the segmentation results in Fig. 5, we notice that, as expected, significantly fewer neurons are dedicated to the representation of background as compared with Fig. 4(d), thereby freeing up neurons to capture foreground regions, where microbial material exists (i.e. regions of high signal activity). The extent of these regions is reduced or 'tightened' as we increase the significance of XY features (a)-(c), whilst the range of intensities captured by each neuron (region) is extended or 'relaxed'. From this it becomes evident that there is an inherent trade-off between signal & proximal resolution as we bias the significance ratio between signal & position.

In the final experiment, we investigate what would happen if there is compromise between features over the course of learning. Using the modifications discussed in Section II.B.4, we show results for the dynamic weight adjustment of XY features, decreasing their role in the

search phase of competitive learning, over time. Fig. 6 shows tests using a bias of 0.8 for GL, GL_{dsk} and PC. In each case (a), (b) and (c), we begin with an initial XY weight of 1, 0.5 and 0.3 respectively, which then decays by a factor of 0.95 over time. In (a), vigorous partitioning occurs in XY whilst it is more significant than the GL features, thus tight regions are captured as in Fig. 4(d). As the significance switches back to signal features however (XY weight falls below 0.8), the remaining neurons are pushed toward capturing signal features. Unlike Fig. 4(d), background regions not having much variation in signal tend not to be further partitioned as the constraint on XY is relaxed, whilst biomass regions (foreground) are. In Fig. 6(b)-(c), we reduce the initial XY weighting such that signal characteristics are favoured from the outset. In decaying this weight though, we induce a similar effect, which achieves some compromise between the results shown in Fig. 5. That is, in both cases, reasonably tight regions are inferred early within the biomass, as in Fig. 5(c). Relaxation then improves detail within biomass regions, as with Fig. 5(a). It is interesting to note that in Fig. 6(b), a slightly higher initial XY weight enables the SOTM to capture more elongated, filamentous structures, whilst the lower initial weighting in Fig. 6(c) favours finer biomass details.

IV. CONCLUSIONS

The flexibility and efficiency of the SOTM to adaptively build a representation of the input feature space has revealed interesting properties via dynamic adjustment of feature significance in the competitive search phase of learning. In this early stage of testing, it is clear that there is an important interplay between positional and graylevel characteristics when building a representation of an 'object' or 'region' in an image with the SOTM. Evidence suggesting that the SOTM can adaptively warp its topology from the traditional means squared sense, into forms that better reflect the types of segmentations typical of the human visual system. Continued investigation and understanding of such properties will be crucial in such applications where segmentation is a difficult task – such as in the extraction of relevant information from microbiological confocal images.

ACKNOWLEDGMENT

The authors (Steven Liss and Ling Guan) wish to acknowledge the financial support from the Natural Sciences and Engineering Council of Canada. Matthew Kyan gratefully acknowledges the APA scholarship from the Australian Research Council. Images of biofilms were provided by Xuetang Zhu, PhD Candidate, Department of Chemical Engineering and Applied Chemistry, University of Toronto, supervised by Steven Liss.

REFERENCES

- [1] Palmer R.J., Sternberg C. "Modern Microscopy in biofilm research: confocal microscopy and other approaches" *Current Opinion in Biotechnology*. June 1999, Vol 10 (3), pp. 263-268
- [2] Liss S. N., "Microbial flocs suspended biofilms" *Encyclopedia of Environ. Microbiology*, G.Sutton (ed) 2002, Wiley & Sons, NY pp. 2000-2012.
- [3] H. Beyenal, Z. Lewandowski and G. Harkin, "Quantifying Biofilm Structure: Facts and Fiction", *Biofouling*, February 2004 Vol 20 (1), pp. 1-23.
- [4] Kong H. S. "The Self-Organising Tree Map, and its Applications in Digital Image Processing". *PhD Thesis*, 1998, University of Sydney, Australia.
- [5] J. Randall, L. Guan, X. Zhang, W. Li., "Investigations of the self-organising tree map", *Proceedings of ICONIP '99*. November, 1999, Perth, Australia. pp. 724-728.
- [6] P. Kovesi, "Image features from phase congruency" *Viere: Journal of Computer Vision Research*. Summer 1999, Vol 1 (3), The MIT press, pp. 1-26.

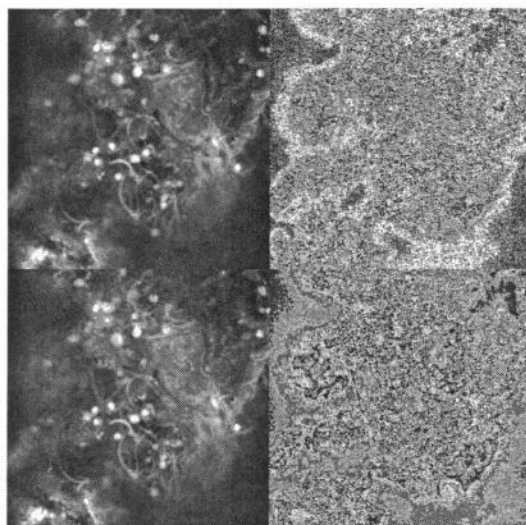


Fig. 4 (a) top: GL only, (b) bottom: GL, GL_{dsk}

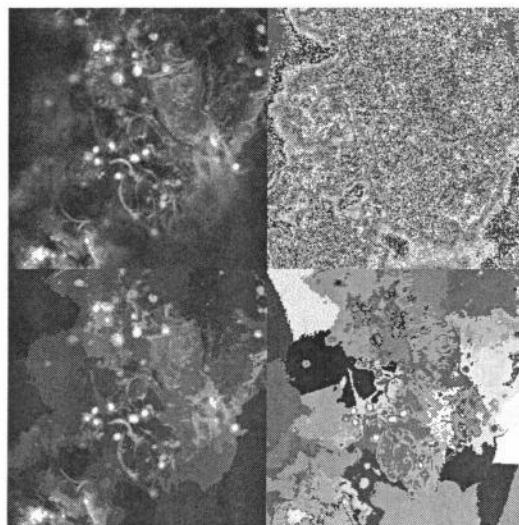


Fig. 4 (c) top: GL, GL_{dsk}, PC (d) bottom: GL, GL_{dsk}, PC, X, Y

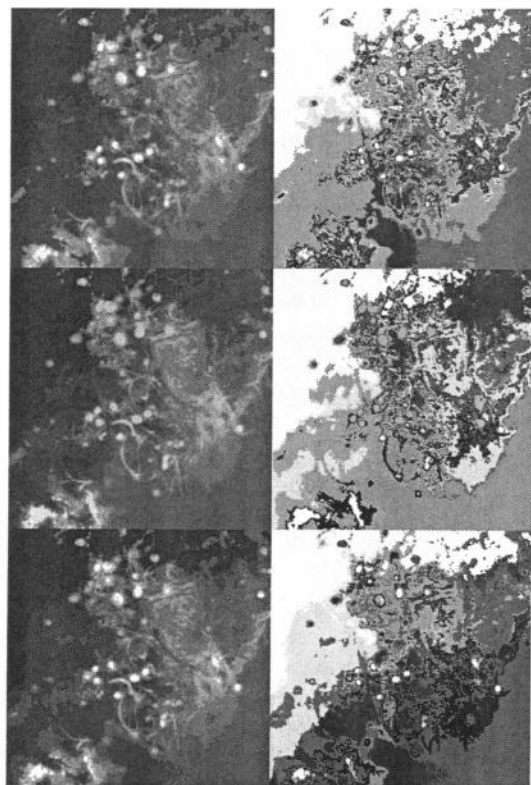


Fig. 5 (a) top, (b) middle, (c) bottom

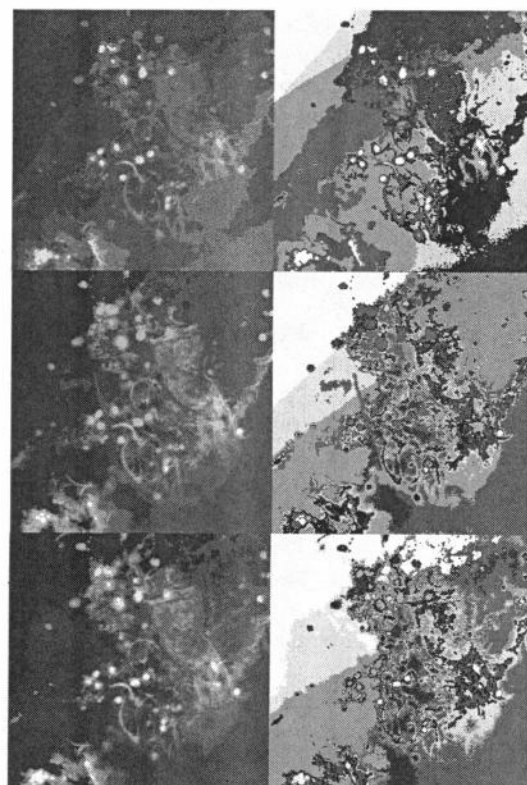


Fig. 6 (a) top, (b) middle, (c) bottom