Bioinformatics Advance Access published April 1, 2004

Extraction of fluorescent cell puncta by adaptive fuzzy segmentation

Tuan Pham[1], Denis Crane[2,3], Tuan Tran[1], and Tam Nguyen[2,3]
[1]School of Computing and Information Technology
[2]Eskitis Institute for Cell and Molecular Therapies
[3]School of Biomolecular and Biomedical Sciences
Griffith University, Nathan Campus, QLD 4111
Australia

ABSTRACT

Motivation: The discrimination and measurement of fluorescently labelled vesicles using microscopic analysis of fixed cells presents a challenge for biologists interested in quantifying the abundance, size and distribution of such vesicles in normal and abnormal cellular situations. In the specific application reported here, we were interested in quantifying changes to the population of a major organelle, the peroxisome, in cells from normal control patients and from patients with a defect in peroxisome biogenesis. In the latter, peroxisomes are present as larger vesicular structures with a more restricted cytoplasmic distribution. Existing image processing methods for extracting fluorescent cell puncta do not provide useful results, and therefore there is a need to develop some new approaches for dealing with such task effectively.

Results: We present an effective implementation of the fuzzy c-means algorithm for extracting puncta (spots), representing fluorescently labelled peroxisomes, which are subject to low contrast. We make use of the quadtree partition to enhance the fuzzy *c*-means based segmentation and to disregard regions which contain no target objects (peroxisomes) in order to minimize considerable time taken by the iterative process of the fuzzy *c*-means algorithm. We finally isolate touching peroxisomes by an aspect-ratio criterion. The proposed approach has been applied to extract peroxisomes contained in several sets of colour images, and the results are superior to those obtained from a number of standard techniques for spot extraction.

Availability: Image data and computer codes written in Matlab are available upon request from the first author.

Contact: t.pham@griffith.edu.au

INTRODUCTION

There are a number of recently developed methods for the analysis of DNA microarray spots such as the morphology-based method proposed by Angulo and Serra (2003); the combinatorial image analysis by Glasbey and Ghazal (2003); and the adaptive thresholding by Liew *et al.* (2003). However, these morphological and statistical thresholding methods are only effective for extracting DNA microarray spots having similar sizes and contained in gridded structures. A main reason for the unsuitable applications of these methods for the segmentation and extraction of biological images in this study is that these images contain very variable (i) spot sizes, (ii) intensity distributions, and (iii) backgrounds. Therefore extraction of these flourescent cell puncta using these methods will lead to over/under-segmented results. An associated method for spot extraction has been developed by Xu *et al.* (1999), which is based on double thresholding and contour-based curve fitting to segment the images of skin cancer. This method is suitable for the segmentation of isolated spots whereas the problem we study herein is not restricted to such cases, and its curve fitting technique can only approximate the spot areas, that may lead to a considerable error for the quantification of peroxisome abundance.

In this paper, we present a segmentation method based on the fuzzy *c*-means for dealing with the more challenging application of extracting and measuring cell puncta images that exhibit low contrast and variable size and cellular distribution, including clustering. The specific application used to test this method is an analysis of the population of peroxisomes in human patient cell lines. Previous findings have indicated a change in the size, cytoplasmic distribution and potential clustering of these cellular organelles in different peroxisomal diseases (Chang *et al*, 1999). We test our proposed segmentation algorithm for extracting cell puncta with real image and compare the

results with those obtained by other standard segmentation methods as well as a current medical image analysis software for spot extraction.

IMPLEMENTATION

Cell culture and immunofluorescence microscopy

Skin fibroblast cell lines were cultured in Dulbeccos Modified Eagle Medium (high glucose), supplemented with 10penicillin-100U/mL streptomycin (Gibco BRL). Cells were processed for indirect immunofluorescence as previously described (Maxwell *et al*, 1999; 2002) and peroxisomes detected using a rabbit antibody to the peroxisomal membrane protein PEX14 and an FITC-labelled goat anti-rabbit secondary antibody (Chemicon). Cells were visualized using a Nikon Eclipse E800 fluorescence microscope equipped with an FITC filter. Images were captured with a Photometrics Coolsnap CCD camera (Roper Scientific) and processed using V++ Precision Digital Imaging software (Digital Optics).

Fuzzy c-means algorithm

The fuzzy c-means (FCM) algorithm (Bezdek, 1981) seeks to partition a dataset $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$, where $\mathbf{x}_m = (x_{m1}, x_{m2}, \ldots, x_{mk})$, $m = 1, 2 \ldots, M$, into a specified number of fuzzy regions which are represented by the corresponding cluster centers. The degrees of each $\mathbf{x}_m$ that belongs to different clusters are characterized by the corresponding fuzzy membership grades taking real values between 0 and 1.

In principle, the FCM maximizes the following objective function:

$$J(\mathbf{U}, \mathbf{c}_1, \ldots, \mathbf{c}_N) = \sum_{y=1}^{N} \sum_{m=1}^{M} \mu_{ym}^{\alpha} \, d_{ym}^2 \qquad (1)$$

where M is the number of data points, N is the number of clusters, $\mathbf{U}$ is the $N \times M$ fuzzy membership matrix, $\mu_{ym} \in [0, 1]$ is the fuzzy membership grade that indicates the degree $\mathbf{x}_m$ belongs to the fuzzy region y, d_{ym} is a distance measure between cluster center $\mathbf{c}_y$ and data point $\mathbf{x}_m$, and $\alpha \in [1, \infty)$ is the fuzzy exponential weight.

The computations of the cluster centers and the partition matrix $\mathbf{U}$ are updated by an iterative procedure which is described as follows.

1. Given the degree of fuzziness α and initial membership matrix $\mathbf{U}$ with random values of $\mu_{ym} \in [0, 1]$ subjected to

$$\sum_{y=1}^{N} \mu_{ym} = 1, \forall m = 1, \ldots, M$$

2. Update initial cluster centers

$$\mathbf{c}_y^{j+1} = \frac{\sum_{m=1}^{M} \mu_{ym}^{\alpha} \mathbf{x}_m}{\sum_{m=1}^{M} \mu_{ym}^{\alpha}} \qquad (2)$$

3. Update fuzzy membership functions

$$\mu_{ym} = \frac{1}{\sum_{z=1}^{N} \left(\frac{d_{ym}}{d_{zm}} \right)^{2/(\alpha-1)}} \qquad (3)$$

where, using the L_2 norm, d_{ym} is given by

$$d_{ym} = ||\mathbf{x}_m - \mathbf{c}_y||_2$$

4. Compute the objective function according to (1). If it converges or its improvement over the previous iteration is below a certain threshold then stop the iterative process. Otherwise, go to step 2.

Estimating the number of clusters

Taking a first look at the image (412×357) as shown in Figure 1, there appears to be two classes to be segmented. These two classes are the background pixels and the peroxisome puncta. If we apply the well-known Otsu's thresholding method (Otsu, 1979) and the FCM, with the number of clusters N=2, to segment the gray image of Figure 1, we obtain Figures 2 and 3 which are the results given by Otsu method and the FCM respectively. It can be seen that both results overestimate the spot sizes and highlight noise and outliers. These are due to the low contrast of the image and particularly the flourescence around the peroxisome spots. We therefore need to add another cluster to represent the flourescent-shadow pixels, i.e., the number of classes is now three instead of two. Because Otsu method only works for gray-scale images with two classes, we now apply the FCM with N=3 and obtain another result as shown in Figure 4. This result shows some improvement over that obtained by the FCM with N=2. However, overestimation of spot areas and touching spots still remain at some extent. We will tackle these problems by a strategy for sharpening the fuzziness of the peroxisome cluster, an aspect-ratio criterion, and quadtree decomposition, which are presented in the following subsections.

Focusing image spots by sharpening fuzzy regions

Based on the concept of a fuzzy set (Zadeh, 1965) and the notion of the Shannon's entropy (Shannon and Weaver, 1948), the measure of fuzziness of a fuzzy set was initially defined by De Luca and Termini (1972) as follows:

1. The fuzziness of $A = 0$ if A is a crisp set, that is, $\mu_A(x) \in \{0, 1\}$, $\forall x \in X$.

Figure 1. Original image A

Figure 2. Segmentation of image A by Otsu's thresholding

Figure 3. Segmentation of image A by FCM with two clusters

Figure 4. Segmentation of image A by FCM with three clusters

Figure 5. Segmentation of image A by sharpening FCM with three clusters

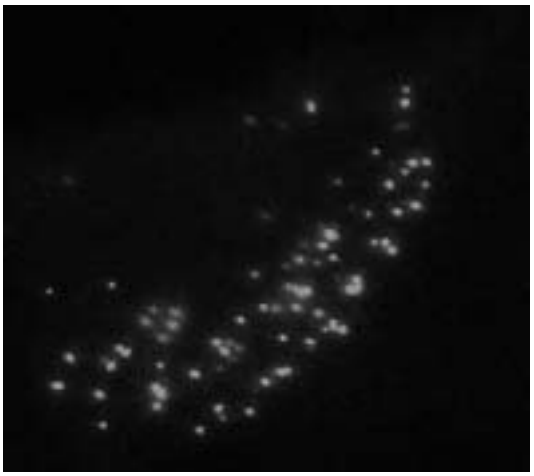

Figure 6. Original image B

2. The fuzziness of A is maximum when $\mu_A(x) = 0.5$, $\forall x \in X$.

3. The fuzziness of A is greater than or equal to that of A^* if A^* is a *sharpened* version of A, that is, $\mu^*_A(x) \geq \mu_A(x)$ if $\mu_A(x) \geq 0.5$; and $\mu^*_A(x) \leq \mu_A(x)$ if $\mu_A(x) \leq 0.5$.

Let $\mu_P(\mathbf{x})$ be the fuzzy membership grade that indicates how possible a pixel $\mathbf{x}$ belongs to the set containing all the peroxisome images, we then apply the notion of the measure of fuzziness to sharpen the fuzzy region of interest (peroxisome) by defining

$$\mu^*_P(\mathbf{x}) = \begin{cases} 1 & : \quad \mu_P(\mathbf{x}) \geq \delta_\mu \\ 0 & : \quad \mu_P(\mathbf{x}) < \delta_\mu \end{cases} \tag{4}$$

where $0.5 < \delta_\mu < 1$ is a fuzzy membership threshold.

What we discuss next is how to get an appropriate value for δ_μ in order to obtain good sharpened peroxisome spots which can make the task of isolating touching spots easier. To fix a concrete idea, let $\mu_{\mathbf{c}^*}(\mathbf{x})$ be the fuzzy membership grade of a pixel $\mathbf{x}$ belonging to the peroxisome cluster $\mathbf{c}^*$. We can say that an optimal value of $\mathbf{c}^*$ must be some value between the *least*, denoted by $f_{\min}(\mathbf{x}|\mathbf{c}^*)$, and the *most*, denoted by $f_{\max}(\mathbf{x}|\mathbf{c}^*)$, bright intensities which are to be assigned to $\mathbf{c}^*$. Of course, it is difficult to determine $f_{\min}(\mathbf{x}|\mathbf{c}^*)$ readily; however, $f_{\max}(\mathbf{x}|\mathbf{c}^*)$ is immediately available, that is, by checking the membership grade of the brightest pixel of the whole image assigned to $\mathbf{c}^*$ given by the FCM. We therefore select $\delta = \mu_{\mathbf{c}^*}(\mathbf{x}^*)$, where $f(\mathbf{x}^*)$ is the maximum intensity value, because $\mu_{\mathbf{c}^*}(\mathbf{x}^*)$ represents the brightest and the least bright pixels which are to be assigned to $\mathbf{c}^*$. Finally, each segmented peroxisome region will be filled up by in case if there are any holes in the region. This is because there exist some low-intensity pixels within the regions. It is also mentioned, as for the problem under study, that the flouresence-processed puncta are always brighter than any other objects (background and noise represented by flourescence stain). And therefore selecting the brightest pixel for δ_μ will not be affected by noise and outliers.

Figure 5 shows an improved segmentation version, in comparison with the result as shown in Figure 4, by applying the sharpening procedure defined in (4) – the segmented spot areas are sharpend and closer to the real spot areas than the former segmented results; in addition, more outliers are removed in this sharpened version.

Isolating touching spots by aspect-ratio criterion

We define an aspect ratio of a spot image p, based on which touching spots can be isolated, as

$$r(p) = \frac{w_{\min}(p)}{w_{\max}(p)}$$

where $w_{\min}(p)$, and $w_{\max}(p)$ are the minimum and maximum widths of the spot area, and $w_{\min}(p) \geq$ the maximum width of the estimated smallest spot size.

The procedure for splitting touching spots is described as follows.

1. Given a spot image p^i, $i = 1, \ldots I$, where I is the number of segmented spots which are greater than an estimated smallest spot image.

2. If $r(p^i) < 0.5$, then split p^i into two subimages p^i_1 and p^i_2 at the location of $w_{\min}(p^i)$.

   (a) If p^i_g, $g = 1, 2$, is greater than an estimated smallest spot size and $r(p^i_g) < 0.5$, then separate p^i_j into two subimages $p^i_{g,1}$ and $p^i_{g,2}$ at the location of $w_{\min}(p^i_g)$.

   (b) Repeat step (a) for all subimages $p^i_{g,\ldots,G}$ where each subscript takes the values from 1 to 2.

3. Repeat steps (1) and (2) for all p^i.

Adaptive segmentation by quadtree partition

What has been described above regarding the fuzzy membership threshold δ_μ exprresed in (4) is a non-adaptive case for the FCM-based segmentation because δ_μ remains the same for the whole image. We notice that, firstly, the regions of interest (peroxisome) occupy only part of the image; secondly, if

we apply the FCM to segment these images with a large size of 1392 by 1040 pixels, the computational time will be considerably long and not so effective for real applications; and thirdly, as an important factor regarding to the parameter δ_μ of which sensitivity depends on the brightest pixel, and if the brightest pixel is not chosen locally then many real spots having relatively low intensities will be subjected to false rejection. We therefore apply the scheme of quadtree partition, that has been largely used for fractal image compression (Fisher, 1994), to iteratively divide the whole image into quadrants so that both segmentation quality and speed will be much enhanced, particularly the second and third issues. By doing this, the segmentation now becomes an adaptive process in which the threshold δ_μ will be estimated differently for each image quadrant.

The image will be partitioned into quadrants (upper left, upper right, lower left, and lower right) if its variance is equal or greater than a splitting threshold δ_{var}, that is

$$var = \frac{1}{N}\sum_{n=1}^{N}[f(x,y) - \bar{f}(x,y)]^2 \ \geq \delta_{var} \qquad (5)$$

where N is the total number of pixels within a (sub)image, $f(x,y)$ and $\bar{f}(x,y)$ are the pixel intensity and the average pixel intensities of the (sub)image respectively.

In order to avoid carrying out the FCM-based segmentation of sub-regions containing all background pixels, we define another decision parameter, denoted as δ_{seg}, based on which the FCM-based segmentation will be performed if the maximum intensity value within a sub-image is greater than a threshold, that is, the decision is to do the fuzzy segmentation if

$$f_{\max}(x,y) \ \geq \delta_{seg} \qquad (6)$$

where $f_{\max}(x,y)$ is the maximum intensity values within a particular sub-image respectively, and δ_{seg} can be experimentally estimated.

Procedure for extracting peroxisome spots

1. Convert the given RBG image into intensity image $\mathcal{I}$.
2. Using a quadtree technique to partition $\mathcal{I}$ into a set of Q subimages: $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_Q$.
3. Do FCM-based segmentation for each $\mathcal{I}_k$, $k = 1, \ldots, K$, where K is the number of quadtree-split images which contain peroxisome spot(s), i.e., $K \leq Q$.
4. Sharpening and fill up spot areas (if there are any holes).
5. Isolating touching spots in each $\mathcal{I}_q$, $q = 1, \ldots, Q$, using the aspect-ratio criterion.
6. Assemble all segmented versions of $\mathcal{I}_q$, $q = 1, \ldots, Q$ to obtain the whole segmented version of $\mathcal{I}$.

RESULTS AND DISCUSSION

In addition to the illustrations, which have been presented in the foregoing sections, showing some advantages of our FCM-based segmentation approach, we further test our proposed method and compare the results with other methods for image spot extraction. For the current FCM analysis, we select $\alpha = 2$, and δ_{seg} to be the rounding off of (255/2) for all cases, for extracting peroxisome image spots on several real images. The reason for choosing the value of $\alpha = 2$ is based on the most popular choice for the FCM analysis found in literature as there is no certain analytical ground for selecting the right value of this paramter at present (Bezdek, 1981; Chi *et al.*, 1996); and for δ_{seg} being half of 255 is based on pre-experiment on a few images from which δ_{seg} was found fairly constant and only large discrepancy on the values of this parameter will turn on or turn off the decision for the FCM analysis.

Figure 6 shows the intensity version of an RGB colour image (412×357) that contains flourescent-stained peroxisome spots. Edges of these spots are fuzzy due to low constrast, also some of the spots are connected to each other. Some flourescent stains may misrepresent spots (false spots) for simple segmentation methods. Figures 7-10 show the segmented versions using Otsu thresholding method, FCM with three clusters, ImageJ that is a public-domain image processing software and can be downloaded from the web (http://rsb.info.nih.gov/ij/), and our proposed FCM-based segmentation method. It can be seen from these figures that the results obtained from both Otsu's thresholding, straight-forward FCM, and ImageJ that uses a thresholding method developed by Ridler and Calvard (1978), show false as well as overestimated peroxisome spots; whereas our proposed method yields the segmentation results that are quite close to the actual spots and can also isolate touching spots.

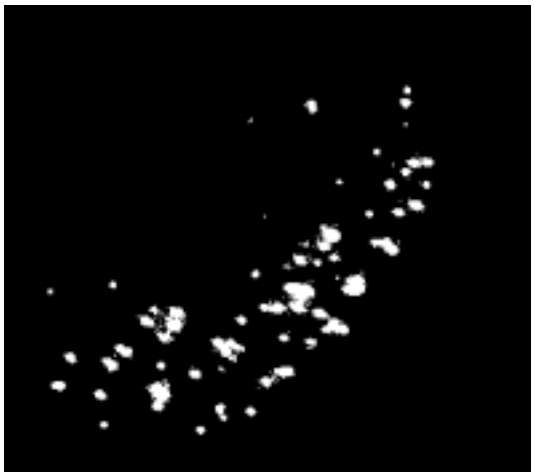

Figure 7. Segmentation of image B by Otsu's thresholding

Figure 8. Segmentation of image B by FCM with three clusters

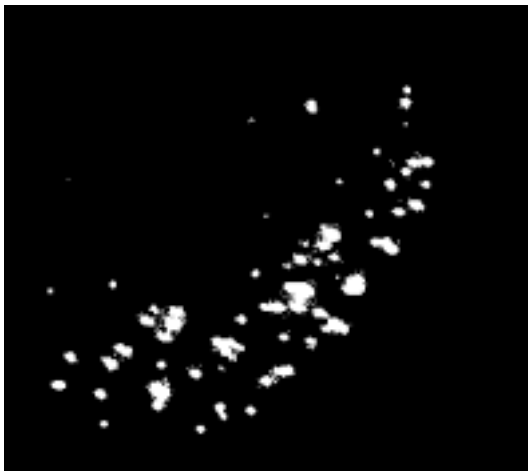

Figure 9. Segmentation of image B by ImageJ (iterative thresholding)

Figure 10. Segmentation of image B by proposed FCM-based method

As another experiment, Figure 11 shows the original image where the task of spot extraction is more difficult than the earlier case in that the image contains many noisy spots. Figures 12-15 show the segmented versions using Otsu's thresholding method, FCM with three clusters, ImageJ, and the proposed FCM-based segmentation. The result obtained by our approach is more accurate than the other three methods. Few small fading peroxisome spots are omitted by our method whereas relatively large number of false spots are detected by the other three methods, particularly by Otsu's thresholding and the ImageJ.

Figures 16-17 shows the Canny edges (Canny, 1986) of the results obtained from ImageJ (Figure 14), and the proposed method (Figure 15) respectively. Again it can be seen that the edges of the peroxisome spots obtained by our method are much more realistic than those of the ImageJ. Spot areas obtained from the ImageJ are significantly overestimated from the actual spot sizes shown in Figure 11; whereas the proposed FCM-based segmenttaion approach yields a more accurate result with the spot areas being close to the actual spots. Touching spots are also isolated by the proposed FCM-based method.

Figures 18-20 show the full-size (1392×1040) versions of the original image, ImageJ-based, and the proposed FCM-based segmentation results respectively. Not only is the proposed method able to yield more accurate spot areas, it is also able to suppress noise and isolate touching spots.

Figures 21-23 show the full-size (1392×1040) versions of the another original image, ImageJ-based, and the proposed FCM-based segmentation results respectively. Conclusion for this case can be drawn

the same as stated above for the results shown in Figures 18-20.

It is mentioned that from all of the above presented results, the extraction of the number of spots and the spot sizes obtained by our method gained the most favour of several biologists at the Eskitis Institute for Cell and Molecular Therapies, and the School of Biomolecular and Biomedical Sciences, Griffith University. From various results, the method is reasonably robust against noise as many low-contrast puncta were detected, isolated and their sizes were more accurately estimated than the other methods.

CONCLUSIONS

We have presented an effective algorithm for extracting flourescent peroxisome puncta in fuzzy images where the contrast is low, spots are touching, and background is mixed with flourescence, which make standard techniques for image segmentation or edge detection ineffective. We have tested our proposed FCM-based algorithm with real image data and obtained favourable results and in all cases have superior results in comparison with existing methods. This algorithm is expected to prove useful for the analysis of different cell compartments following fluorescence microscopy.

Figure 11. Original image C

Figure 12. Segmentation of image C by Otsu's thresholding

Figure 13. Segmentation of image C by FCM with three clusters

Figure 14. Segmentation of image C by ImageJ (iterative thresholding)

Figure 16. Canny-edge image of segmentation by ImageJ (iterative thresholding)

Figure 15. Segmentation of image C by proposed FCM-based method

Figure 17. Canny-edge image of segmentation by proposed FCM-based method

Figure 18. Original image D

Figure 19. Segmentation of image D by ImageJ (iterative thresholding)

Figure 20. Segmentation of image D by proposed FCM-based method

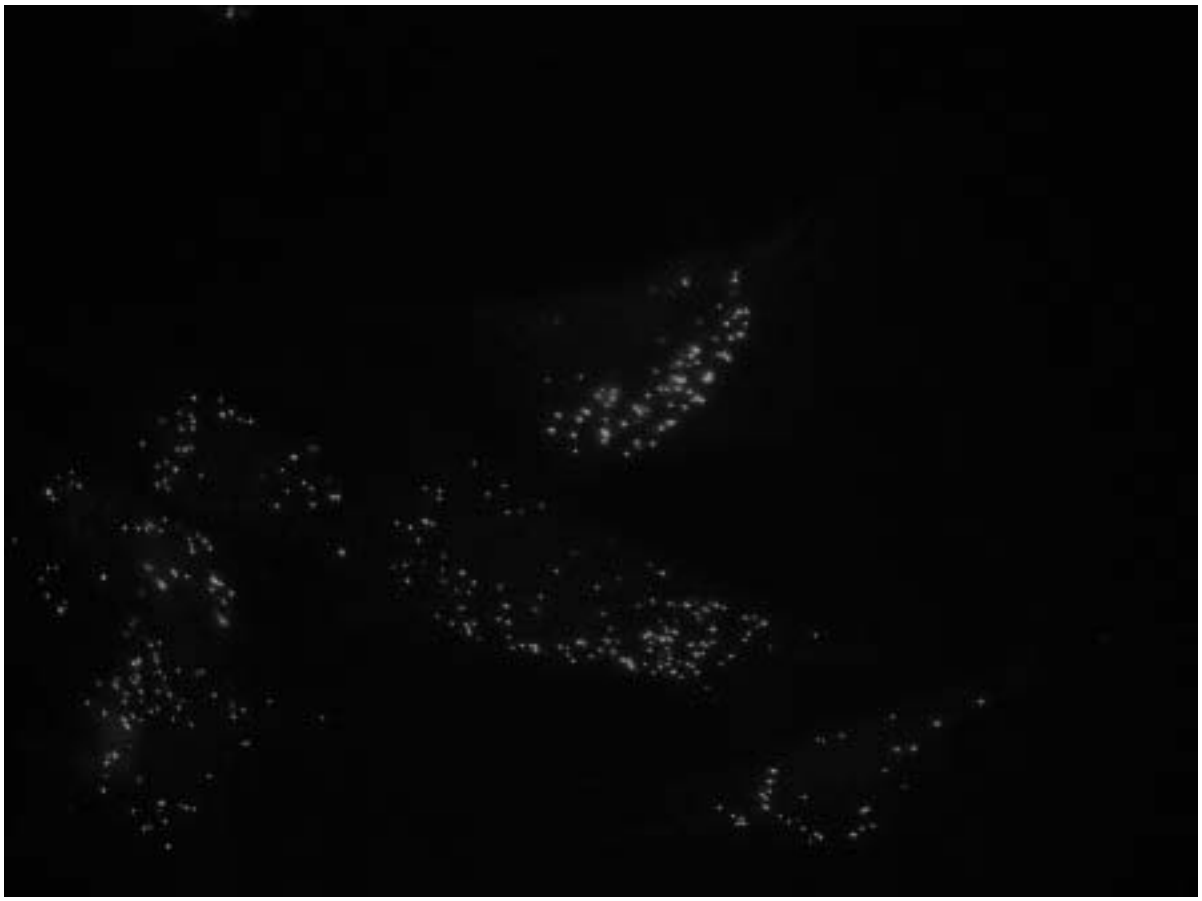

Figure 21. Original image E

Figure 22. Segmentation of image E by ImageJ (iterative thresholding)

Figure 23. Segmentation of image E by proposed FCM-based method

REFERENCES

Angulo, J., and Serra, J. (2003) Automatic analysis of DNA microarray images using mathematical morphology. *Bioinformatics*, **19**, 553-562.

Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.

Canny, J. (1986) A computational approach for edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 679-698.

Chi, Z., Yan, H., and Pham, T. (1996) *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*. World Scientific Publishing, Singapore.

Chang, C. C., South, S., Warren, D., Jones, J., Moser, A.B., and Moser, H.W. (1999) Metabolic control of peroxisome abundance. *Journal of Cell Science*, **112**, 1579-1590.

DeLuca, A. and Termini, S. (1972) A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, **20**, 301-312.

Dembele, D., and Kastner, P. (2003) Fuzzy c-means method for clustering microarray data. *Bioinformatics*, **19**, 973-980.

Fisher, Y., Editor (1994) *Fractal Image Compression Theory and Application*. Springer, New York.

Glasbey, C.A., and Ghazal, P. (2003) Combinatorial image analysis of DNA microarray features. *Bioinformatics*, **19**, 194-203.

Liew, A.W.C., Yan, H., and Yang, M. (2003) Robust adaptive spot segmentation of DNA microarray images. *Pattern Recognition*, **36**, 1251-1254.

Maxwell, M.A., Nelson, P.V., Chin, S.J., Paton, B.C., Carey, W.F., Crane, D.I. (1999) A common PEX1 frameshift mutation in patients with disorders of peroxisome biogenesis correlates with the severe Zellweger syndrome phenotype. *Human Genetics*, **105**, 38-44.

Maxwell, M.A., Allen, T., Solly, P.B, Svingen, T., Paton, B.C., and Crane, D.I. (2002) Novel PEX1 mutations and genotype-phenotype correlations in Australasian peroxisome biogenesis disorder patients. *Human Mutation*, **20**, 342-351.

Otsu, N. (1979) A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man, and Cybernetics*, **9**, 62-66.

Ridler, T.W. and Calvard, S. (1978) Picture thresholding using an iterative selection method. *IEEE Trans. Systems, Man, and Cybernetics*, **8**, 630-632.

Shannon, C., and Weaver, W. (1948) The mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379-423, 623-656.

Xu, L., Jackowski, M., Goshtasby, Roseman, D., Bines, S., Yu, C., Dhawan, A., and Huntley, A. (1999) Segmentaion of skin cancer images. *Image and Vision Computing*, **17**, 65-74.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338-353.