Pergamon

Pattern Recognition, Vol. 28, No. 2, pp. 269–281, 1995
Elsevier Science Ltd

0031 3203/95 $9.50 + .00

0031-3203(94)00089-1

WATER QUALITY ANALYSIS: A PATTERN RECOGNITION APPROACH†

DIPTI PRASAD MUKHERJEE, AMITA PAL, S. ESWARA SARMA and D. DUTTA MAJUMDER

National Centre for Knowledge Based Computing, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India

(*Received* 24 *November* 1993; *received for publication* 5 *August* 1994)

Abstract—The development of a microcomputer-based system for the automatic counting of colonies of indicator organisms (that is, bacteria whose presence is indicative of the presence of pathogens) trapped on membrane filters from water samples is described here. Three different approaches, based on mathematical morphology, distance transform and fuzzy *c*-means clustering respectively, are implemented and their relative performances are discussed here.

Coliforms Gray-level opening Distance transform Fuzzy *c*-means clustering

1. INTRODUCTION

The quality of water used by communities is an important concern of users as well as administrators, since the general health of the community is greatly dependent on it. Pollution of water can be chemical or bacterial, this paper being concerned only with the latter. It is well known that many of the serious diseases that afflict humans, like typhoid, cholera and gastroenteritis, are water-borne, so it is important that drinking water should be screened for pathogenic bacteria, on a routine basis, to prevent outbreaks of such diseases.

A certain class of bacteria, called coliforms are reliable indicator organisms for testing water quality because they travel with disease-producing microorganisms and are easy to isolate. Their presence in drinking water usually indicates that water is unsuited for drinking and may need treatment with a disinfectant. In a study,[1] the World Health Organization (WHO) has proposed that the number of *Escherichia coli*, which is a coliform, could be a parameter for determining the quality of drinking water or coastal waters.

Fast and reliable water quality tests are in great demand in view of the increasing need for reusing available water supplies. Most bacteriological water quality criteria are based on examining water samples for faecal pollution indicator bacteria, in particular, total and faecal coliforms. It is essential to monitor the quality of water at all times, so as to control the incidence of water-borne diseases. For this purpose a *fast*, *accurate* and *low-cost* water quality analyzer is very much needed.

† The work has been partially supported by DoE/UNDP (IND/85/072).

The membrane filter technique[2] is a relatively new and reliable laboratory technique for the bacteriological examination of water. Basically, the idea is to pass a known volume of the water sample under examination through a membrane filter, whose pore size is small enough (0.45 μm) not to allow *E. coli* and other coliforms to pass through. When this filter is incubated on an appropriate medium at a suitable temperature and for the right length of time (generally about 24 h), colonies develop upon the filter disk wherever bacteria were entrapped during the filtration process. Depending upon the medium used, the colonies of *E. coli* take on a colour which is different from those of other bacteria, so that they can be identified easily and counted, using either a low-power microscope or a magnifying glass. Membrance filters may have grid lines etched on them to facilitate the counting process.

A flow-chart for the MF technique is given in Fig. 1. The counting process is normally carried out manually by laboratory technicians, so that the processing of large numbers of water samples is rather tedious and prone to error. Also, laboratory technicians generally do not take a count of all the colonies on a filter. What is normally done is that a few squares are selected using some rule, and the average count for these squares is multiplied by the total number of squares in the grid, to give a reasonable estimate of the total number. Therefore, the authors have made an attempt to develop a microcomputer-based system that can automate the counting, covering all the colonies on the filter in the process. Basically the steps involved are as follows:

(1) capture of the image of the membrane filter by means of a CCD camera;

(2) preprocessing of the image by the use of appropriate image processing techniques;

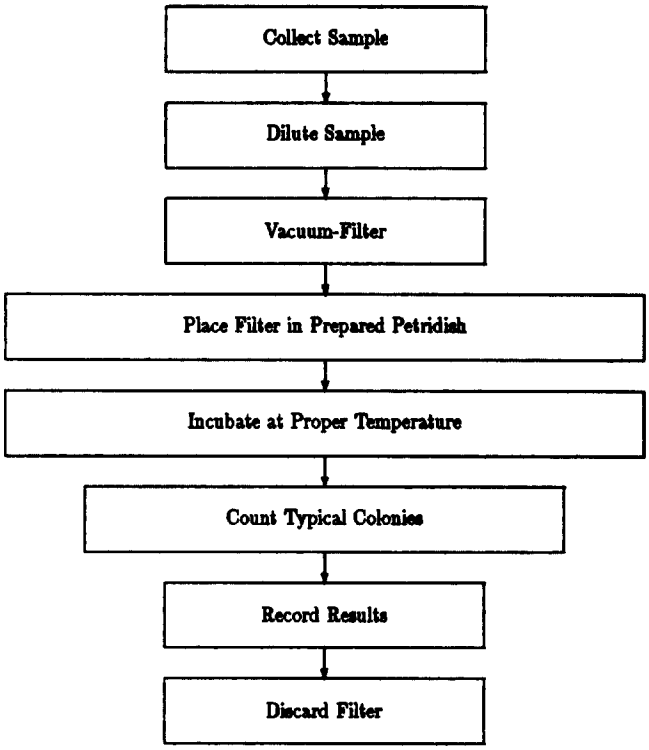


Fig. 1. Flow-chart for the MF technique.

(3) development of algorithms for the classification and counting of the coliform colonies.

A set of techniques for separating and counting the colonies are described below. Our intention is to present to the user industry a set of methods which will be robust and user-friendly, the criteria for judging the performance being:

- the processing time,
- the accuracy of the method, and
- the degree of user interaction.

So far, we have tried to combine simple conventional techniques and obtained reasonably good results that require little processing time and minimal user interaction.

The three methods that we had tested so far for colony segmentation from the background and grid lines (on the filter) are based respectively on:

(1) gray-level morphology,[3]
(2) distance transform,[4] and
(3) fuzzy c-means classification approach.[5]

In all the above cases, except for classification, the counting is done by implementing the component-labeling algorithm[6] on the segmented image. The size of a single colony is determined from the size distribution of all the colonies obtained using component labeling. The mode of the distribution gives the size of the single colony as we are assuming that the frequency of occurrence of single *E. coli* colonies far exceeds those for overlapping colonies. We describe below each method separately followed by the results obtained. Finally, we present a comparative study of different methods and state our recommendations to the user industry. We conclude with the issues we are currently concentrating on, and on future challenges.

2. OUR APPROACH

2.1. *Problem definition*

Broadly speaking, the problem is to identify and count bacterial colonies from images. This has been reformulated by us as one of classifying pixels into one of three classes—*colony*, *background* and *grid*, the grid being etched on the membrane filter by the manufacturers to aid the lab technicians in the counting process.

Complications arise on account of the presence of the grid on the membrane filter. While the grid serves as an aid for quick counting of colonies for the human operator, in our case, that is, when using Image Processing/Pattern Recognition (IP/PR) techniques, it poses problems, if we treat it (the grid) as a separate class, in addition to *background* or *colonies*. At present, the grid cannot be eliminated, since we require it to prove and calibrate the performance of our algorithm, as compared to the manual counting. This is discussed in subsequent sections.

2.2. *Preprocessing*

At least eight different views of each petri dish are taken under varying front-lighting conditions. This is

done to reduce the effect of specularity from colonies and non-uniform lighting over the viewing surface. Images for all the plates are captured under a constant magnification. We have found the following steps to work satisfactorily.

(i) The average pixel values of the eight views are calculated.

(ii) The sum of the absolute difference of the average from the individual slides scaled linearly from 0 to 255 scale gives a satisfactory enhanced image for colony identification.

However, we are in the process of designing a proper set-up for capturing the images of colonies with both back- and front-lighting, so that non-uniformity of lighting over the entire area is somewhat reduced. We intend to follow.[7] We hope that this step can be eliminated once we have our specialized set-up for capturing images of colonies.

2.3. *Using gray-level morphology*

The algorithmic steps involved in this process are:

(1) gray-level opening of the preprocessed image,
(2) thresholding of the image, followed by
(3) application of the connected component labeling algorithm.

The gray-level opening of a set X by some structuring element Y is defined in terms of an *erosion* followed by a *dilation*.[3]

$$X_Y = [X \ominus Y] \oplus Y. \tag{1}$$

The opening of X by Y is the union of translations of Y completely contained in X. The *dilation* of a set X by a structuring element Y can be expressed

$$X \oplus Y = \bigcup_{y \in Y} X_y \tag{2}$$

while the *erosion* can be expressed as

$$X \ominus Y = \bigcap_{y \in Y} X_{-y}. \tag{3}$$

The dilation of $U[a]$, the umbra representation of a gray-scale image A, by B is the union of translations of $U[a]$ by the points of B. At a given location (x, y), the gray-levels $d(x, y)$ of the dilation are determined by the maxima of the translated umbrae $U[a]$. Therefore, the dilated gray-scale image can be computed as

$$d(x, y) = \max_{i,j} [a(x - i, y - j) + b(i, j)], \tag{4}$$

where $b(x, y)$ describes the surface of B. Similarly, the erosion is determined as the difference of gray-levels of $b(x, y)$ and the points of $U[a]$. The gray-levels for pixels of the eroded image are given by

$$e(x, y) = \min_{i,j} [a(x - i, y - j) - b(-i, -j)]. \tag{5}$$

Therefore, opening a gray-scale image amounts to creating a gap wherever sharp peaks and ridges of gray-value are present.

Due to the differences in gray-values among the pixels in the three classes, namely, *grid lines*, *colonies* and the *background*, the preprocessed images are subjected to gray-level opening. Various mask sizes are tested. The 7×7 mask of $Y[i][j] = 1\ \forall i, j$ is found to be suitable for 512×512 images. A thresholding of the gray-level opened image segments the colonies from the background. The gray-level ridges due to grid lines are first blurred due to opening and subsequently eliminated due to thresholding. We have selected the threshold value by gray-level opening the image of background only. A portion of the image of filter paper is selected without grid line and coliform and the histogram of its gray-value after opening gives the threshold value.

The segmented image is then subjected to connected component labeling algorithm,[6] which is as follows:

(1) all the dark pixels (corresponding to colonies) are marked with integers (starting from 1) in increasing order. The image is scanned in the increasing order of rows, starting from its top-left corner,

(2) in the subsequent iterations, each pixel and its marked value is replaced with the local maxima in an 8-connected neighborhood. The iterations continue till there is no further replacement for all the dark pixels,

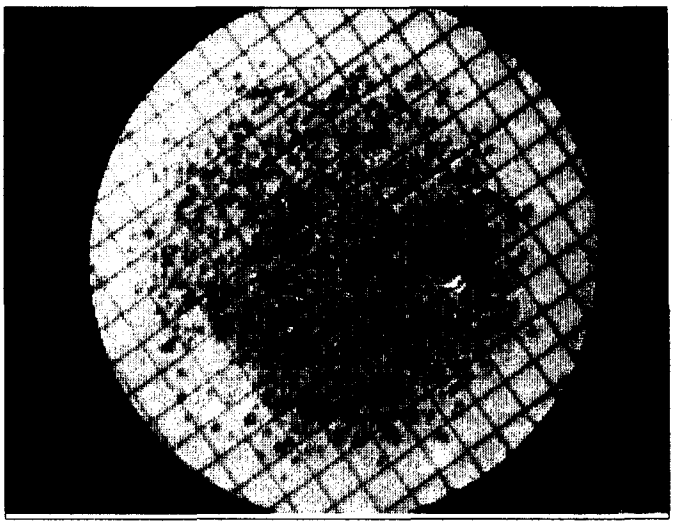
(3) count the occurrences of each pixel marking values from its frequency distribution, and

(4) the frequency distribution of the count in Step 3 gives the number of colonies in the given image. The mode of the distribution gives the size of a single colony after the specified duration of incubation, since it appears logical to assume that single colonies appear in far greater numbers than overlapping colonies. The number of colonies present is calculated by dividing the total number of dark pixels by the mode of the distribution.

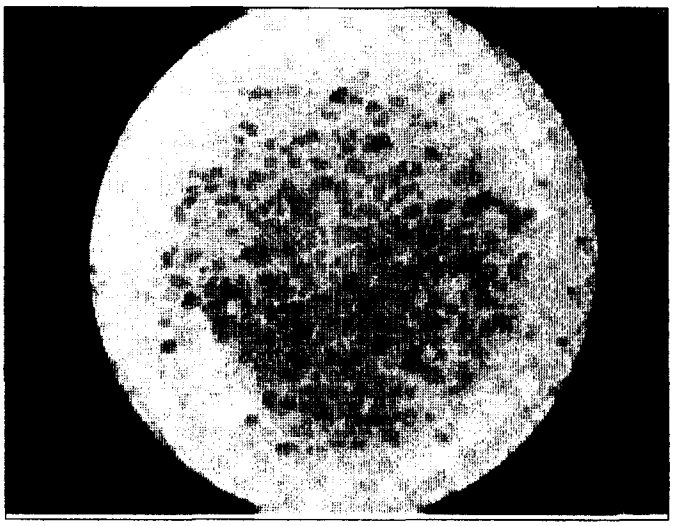
The disadvantages those result from morphological approach, and suggested improvements, are discussed in greater length in Section 3.

2.3.1. *Results.* Figure 2(a) is the reduced image of a membrane filter on which colonies have formed. Figure 2(b) is the output after gray-level opening while the result of gray-level thresholding on Fig. 2(b) is shown in Fig. 2(c). A sub-image of the image in Fig. 2(c) is shown in Fig. 2(d), on which the connected component labeling algorithm is run. Fig. 2(e) is the output after the labeling of connected components while Fig. 2(f) shows the frequency distribution of size of connected components in the image in Fig. 2(e). The mode of distribution is at 25 and the number of colonies present is 430.

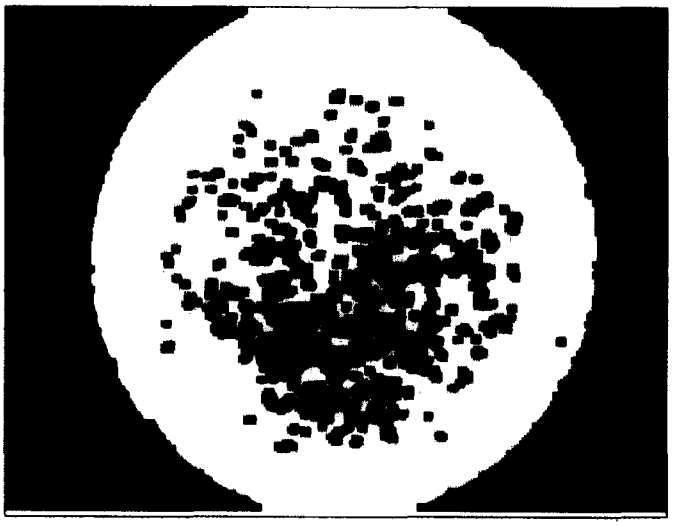
Interested readers may compare the result of Fig. 2(d) with the result obtained in Subsection 2.4, particularly with Fig. 3(d), which is obtained using the distance transform method. Note that in both the cases, the original pictures are the same.



(a)

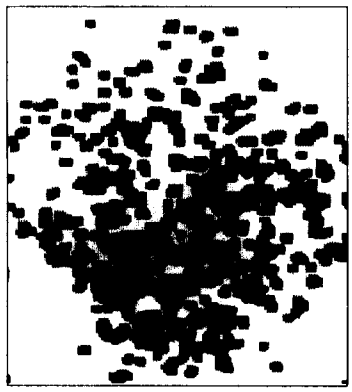


(b)

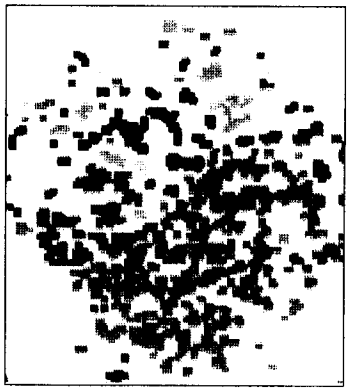


(c)

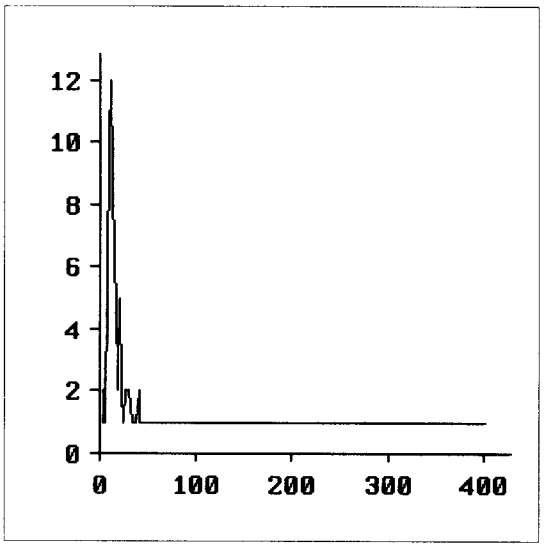
Fig. 2. (a) Original Image. (b) Image in (a) after gray level opening with 7 × 7 masks. (c) Image in (b) after thresholding (at gray value 113). (d) A sub-image of image in (c). (e) Image in (d) after component labeling. (f) Frequency distribution of size of components of image in (e).



(d)



(e)



(f)

Fig. 2 (*Continued*)

2.4. *Using distance transform*

The steps involved in making colony counts, based on the distance transform technique, are as follows:

(1) thresholding of the preprocessed image,
(2) application of the distance transform to the image,
(3) thresholding of the distance-transformed space, the local maxima obtained in 3 × 3 neighborhood being grown to get back the *almost* equivalent size of the colony or *congruence* of colony. The grid lines are eliminated due to thresholding in the distance transformed space, and
(4) application of the connected component labeling algorithm to count the number of colonies (see Subsection 2.3).

The gray-level thresholding is straightforward, the only consideration being that there is no loss of colony components in the thresholded image. Average gray-value of a portion of filter paper image without grid line and coliform organisms determines the threshold limit. The binary image has pixel values 0 and 1, 1 representing colony or grid, while 0 is background. The distance transform of the thresholded image can be achieved in two passes[4] only. For a 3 × 3 mask

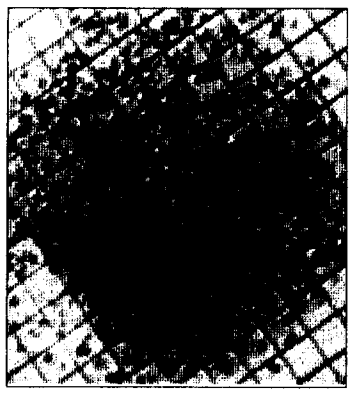
$$\begin{matrix} A & B & C \\ D & E & F \\ G & H & I \end{matrix}$$

in the first pass, starting from the top-left corner of the segmented image, the central pixel E is replaced by

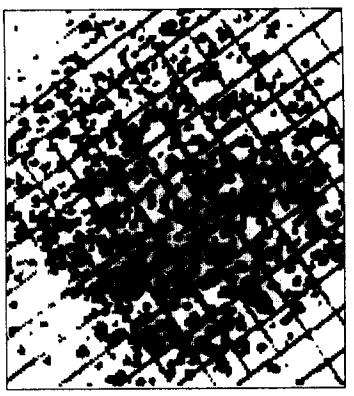
$$E_{\text{new}} \leftarrow \min(D, A, B, C) + 1. \tag{6}$$

In the second pass, originating from the bottom-right corner, the central pixel E is replaced by

$$E_{\text{new}} \leftarrow \min(G, H, I, F, (E - 1)) + 1. \tag{7}$$

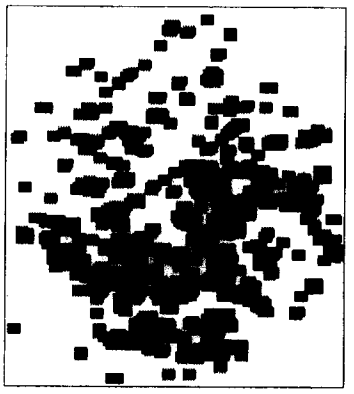


(a)

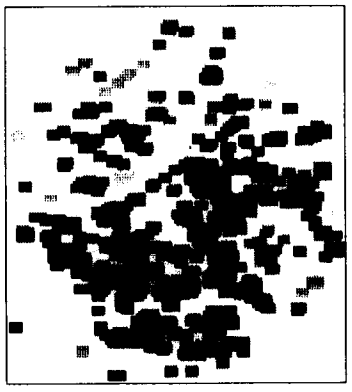


(b)

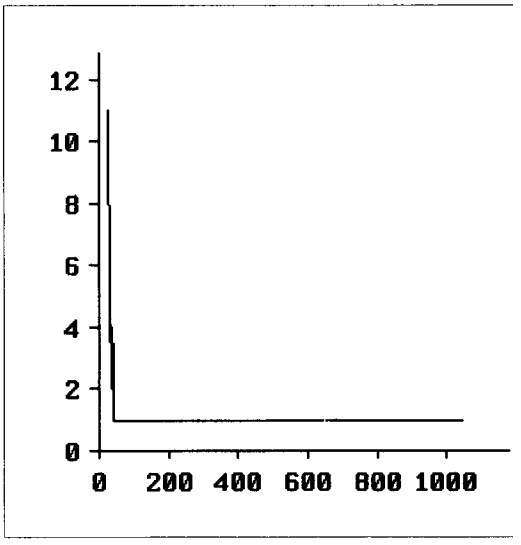
(c)



(d)



(e)



(f)

Fig. 3. (a) A sub-image of image in Fig. 2(a) of identical size of that in Fig. 2(d). (b) Image in (a) after thresholding (at gray value 125). (c) Image in (b) after distance transform. (d) Image in (c) after region growing. (e) Image in (d) after component labeling. (f) Frequency distribution of size of component of the image in (e). Mode of distribution is approximately at 43 and the number of *E. coli* is 398.

The local maxima in the distance-transformed space gives the central or *core* point of each colony or its *congruence*. The central points for the grid lines in the transformed space are eliminated by thresholding. Since, the thickness of the grid line is constant through out the image space, the thresholding is straightforward.

The core points are then expanded to a $(2i-1) \times (2i-1)$ square, where i is the value of the core point which remains in the centre of the square. Note that while expanding, the grown region may contain other core points. A linked list data structure is implemented to keep tag on the connected components. For a set S containing an unmarked core point p_i, iterate the following steps until $S = \varnothing$.

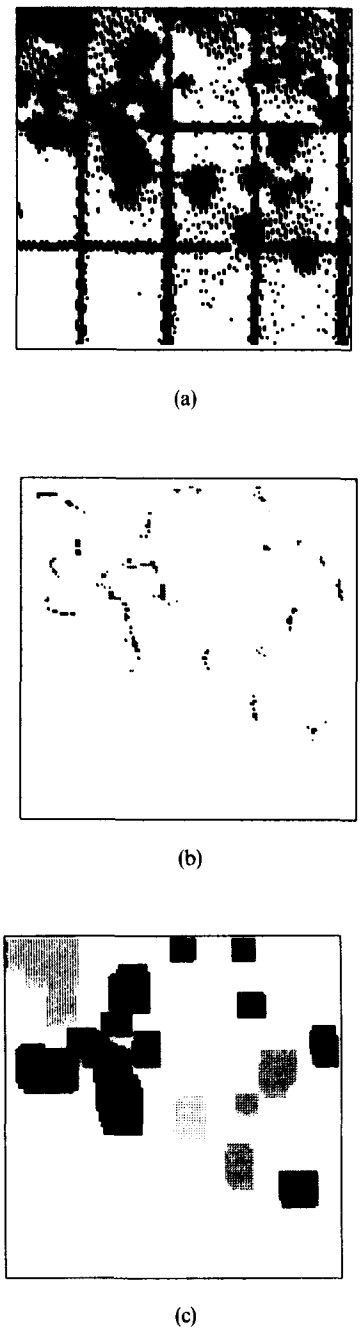
Step 1. For core point p_i with distance transformed value i, mark the square $(2i-1) \times (2i-1)$.
Step 2. Eliminate p_i from set S.
Step 3. Add the core points $p_{i1}, \ldots, p_{iN}$ to set S contained in the square $(2i-1) \times (2i-1)$, excepting the already eliminated core point p_i.

If $S = \varnothing$, all the marked pixels due to iterations of the above steps represent a single connected component. The above region growing procedure will continue with next unmarked p_i until all core points are marked. Due to this region growing, approximate colony structure can be recovered though the exact topology will be somewhat different due to digital analysis of the colony shape.

This region-grown image is then used as input to connected component labeling algorithm as described in Subsection 2.3.

2.4.1. *Results.* We test the same image as in Fig. 2(a), cut in size equivalent to the dimensions of the image in Fig. 2(d), as shown in Fig. 3(a). Figure 3(b) gives the image after thresholding, while Fig. 3(c) shows the local maxima of distance-transformed space after thresholding. Figure 3(d) shows the effect of region growing as described above, maintaining the connected component topology. Figure 3(e) shows the output after the labeling of all the connected components, while Fig. 3(f) gives the distribution of colony size. The mode is at 43 and the number of colonies is 398. The result agrees with those obtained using gray-level morphology as shown in Fig. 2(e). Several such comparisons are carried out and the algorithm using distance transform seems to give results close to $\pm 10\%$ of the algorithm using gray-level morphology.

Figures 4(a) and 5(a) are segmented images with front- and back-lighting, whose original images are Figs 6(a) and 7(a), respectively. Note that there is no grid present in case of back-lighting [Fig. 7(a)]. Figures 4(b) and 5(b) give the local maxima of the distance-transformed space duly thresholded after eliminating core points of the grid lines, particularly for front-lit images. Figures 4(c) and 5(c) give the result of connected component labeling. The counting results are described in the Table 1. Readers can compare the results of Figs 4(c) and 5(c) with Figs 6(b) and 7(b) which are obtained using fuzzy c-means classification algorithm which we will present in the next subsection.



(a)

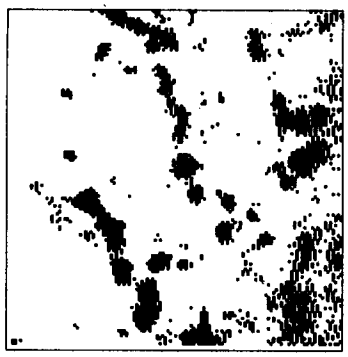
(b)

(c)

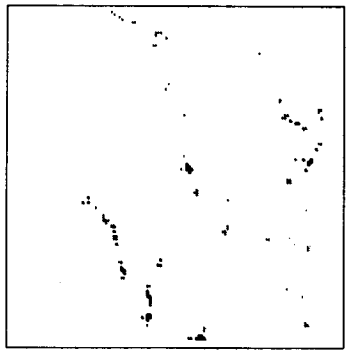
Fig. 4. (a) The image in Fig. 6(a) after thresholding (at gray value 115). (b) The image in (a) after thresholding and finding local maxima of distance transformed space. (c) The image in (b) after region-growing and component labeling.

2.5. *Using fuzzy c-means clustering*

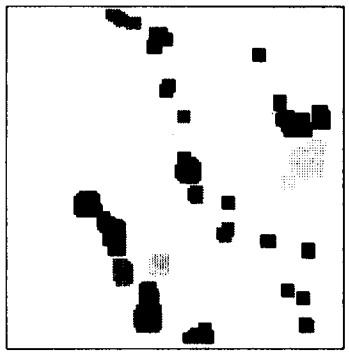
As mentioned earlier, we have formulated the colony counting problem as a classification problem. The three distinct classes present in the slides are the *colonies*,



(a)



(b)



(c)

Fig. 5. (a) The image in Fig. 7(a) after thresholding (at gray value 200). (b) The image in (a) after thresholding and computation of local maxima in distance transformed space. (c) The image in (b) after region-growing and component labeling.

the *grid* and the *background*. The features associated with each pixels are:

(1) the average of gray-values in a 3 × 3 mask, and
(2) the busyness value.

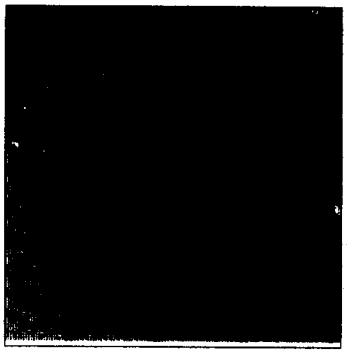
The busyness[5] of each pixel is defined as $\min(v_x, v_y)$ where

$$v_x = |A-B| + |B-C| + |D-E| + |E-F| + |G-H| + |H-I| \quad (8)$$

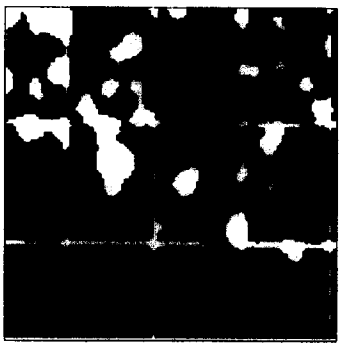
Table 1. Overall result

Figure no.	Coliform count	Details
2(d)	430	Using gray-level morphology
3(a)	398	Using distance transform
4(a)	42	Using distance transform
5(a)	37	Using distance transform (B)
6(a)	72	Using fuzzy *c*-means clustering c = 3, v = 2, m = 2
7(a)	42	Using fuzzy *c*-means clustering (B) c = 3, v = 2, m = 2
8(a)	305	Using fuzzy *c*-means clustering c = 3, v = 2, m = 2
9(a)	39	Using fuzzy *c*-means clustering after detecting linear structure using Hough transform c = 3, v = 2, m = 2

c = class; v = feature vector; m = fuzzy exponent; B = backlit.



(a)



(b)

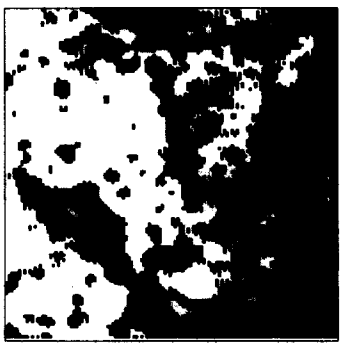
Fig. 6. (a) Original Image. (b) The image in (a) after fuzzy *c*-means clustering with $c = 3$.

$$v_y = |A-D| + |D-G| + |B-E| + |E-H| + |C-F| + |F-I| \quad (9)$$

and

$$\begin{matrix} A & B & C \\ D & E & F \\ G & H & I \end{matrix}$$

(a)



(b)

Fig. 7. (a) Backlit original image. (b) The image in (a) after fuzzy c-means clustering with $c = 3$.

is the 3 × 3 mask where the busyness value is calculated for the pixel E. Instead of $\min(v_x, v_y)$, we found that the busyness value equivalent to $\max(v_x, v_y)$ for the three classes mentioned above has more discriminating power. An additional feature, namely, the average of (v_x, v_y) is also tested, though the result with three features are not found to be satisfactory.

We report here the application of fuzzy c-means algorithm[5] to discriminate the bacterial colonies from the grid and the background. The clustering algorithm is based on a least-squared error criterion. The squared error clustering criterion is given by

$$\mathbf{J}_m(\mathrm{U}, \mathrm{v}) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mathbf{u}_{ik})^m (d_{ik})^2,$$

where U belongs to Fuzzy c-partition space and $\mathbf{u}_{ik}$ contains the fuzzy membership value of kth data point to ith class, v are the cluster centers of fuzzy c-partitions. These are updated in every iteration based on $\mathbf{u}_{ik}$ and the distance $d_{ik} = \|\mathbf{X}_k - \mathrm{v}_i\|$, between ith cluster centre of v and kth data point of $\mathbf{X}$. This continues till the cluster centres become stable and there is *insignificant* difference between cluster centres in two consecutive iterations. m is the weighting exponent where $m \in [1, \infty)$.

To *minimize* $\mathbf{J}_m(\mathrm{U}, \mathrm{v})$ subject to the condition $\sum_{i=1}^{c} \mathbf{u}_{ik} = 1, 0 < \sum_k \mathbf{u}_{ik} < k$ and $\mathbf{u}_{ik} \geq 0$, applying Lagrange multipliers to the variables $\mathbf{u}_{ik}$, we obtain

$$\mathbf{u}_{ik} = \frac{1}{\left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}\right]} \tag{10}$$

and $\forall i$

$$V_i = \frac{\sum_{k=1}^{n} (\mathbf{u}_{ik})^m \mathbf{X}_k}{\sum_{k=1}^{n} (\mathbf{u}_{ik})^m}. \tag{11}$$

In every iteration, $\mathbf{u}_{ik}$ is calculated using equation (10) while v_i is updated following equation (11) using the earlier $\mathbf{u}_{ik}$ value.

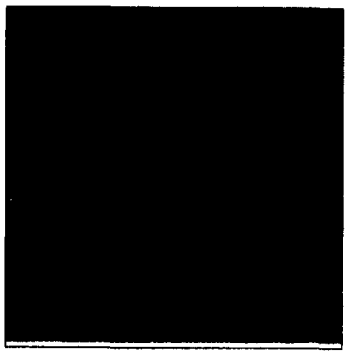
2.5.1. *Training module.* The fuzzy c-means clustering algorithm is applied to a chunk of preprocessed image. The images with all the clusters present in significant proportions are chosen. We find that the cluster seed points stabilizes after 18 to 20 iterations. The output of v_{in} gives seed point for ith class and nth feature vector for further classification while the $\mathbf{u}_{ik}$ gives the fuzzy membership value of the kth data point associated with the ith class. The cluster centres for the *colony* class are detected by visual inspection of the original and classified images, and the gray-values associated with them.

2.5.2. *Classification module.* The pixels are classified based on the highest membership value in $\mathbf{u}_{ik}$ associated with the particular class. To reduce processing time for a large image, pixels are classified based on the minimum distance d_{ik} between the feature vectors for the pixel and the seed points of the classes.

2.5.3. *Results.* Figure 8(b) is the output of preprocessing step on Fig. 8(a). Figure 8(c) is the classified image based on feature vector consisting of gray-value averaging and busyness. The frequency distribution of gray-value averaging and busyness are shown in Figs 8(d) and 8(e), respectively.

Figures 6(b) and 7(b) are the classified images of the original images Figs 6(a) and 7(a), respectively. We found that the classification algorithm gives the most stable results with three classes and two features. The results are repeated with fuzzy weighting exponent values 1.5 and 2.0. However, not much difference is observed in the results. The grid class, in the case of Fig. 6(b) is misclassified to some extent, both to colony and background classes. Figure 7(b) is a back-lit image without grid. Incidentally, in the case of a backlit image, the background itself is found to consist of two different (sub-) classes, possibly due to uneven distribution of nutrient medium on the petridish.

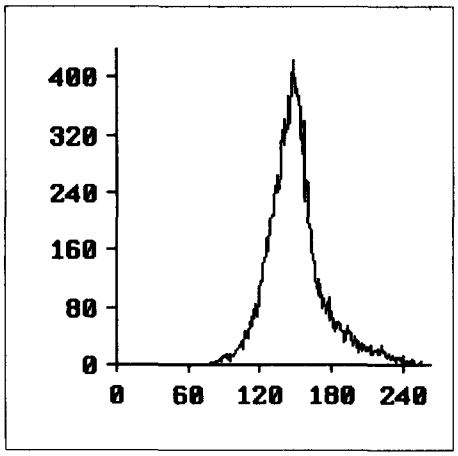
Figure 9(a) is obtained applying Hough transform to Fig. 4(a). It is to be noted that considerable amount of postprocessing is required to find the linear structures in Fig. 4(a). The image is transformed to (r, θ) space and local maxima are calculated after thresholding.



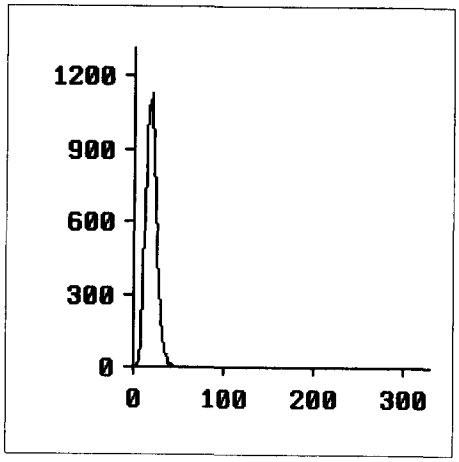
(a)

(b)

(c)



(d)



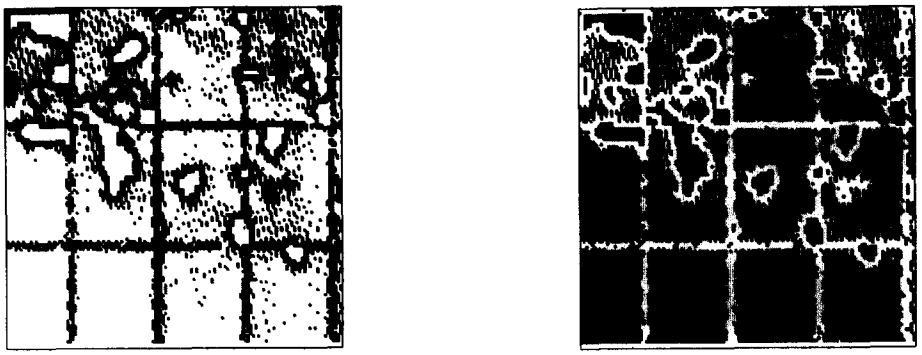
(e)

Fig. 8. (a) Original Image. (b) Image in (a) after enhancement. (c) Image in (b) after fuzzy c-means clustering with fuzzy exponent value $m = 2$. (d) Frequency distribution of gray value averaging using 3×3 mask on the image in (b). (e) Frequency distribution of gray value busyness using 3×3 mask on the image in (b).

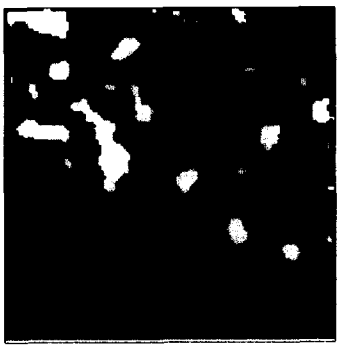
Since, the local maxima value in Hough space due to the colonies and the linear structure (grid lines) are close, it is almost impossible to isolate the grid lines completely from the blob-shaped colonies. Therefore, after the reverse transformation of peaks of Hough space to image space, each potential pixel and its neighborhood in a 5×5 window are compared with the original image. If all the pixels in the 5×5 window found to have gray-value close to those of the colony pixels (i.e. greater than the threshold), then the potential pixel is discarded as being a part of the colony class. Figure 9(a), thus obtained, is then subtracted from Fig. 6(a) to get Fig. 9(b) with grid lines having distinctly separate gray-values. Though the classification problem is quite straightforward now, this process of linear structure detection followed by postprocessing is highly CPU-intensive. We have applied fuzzy c-means clustering algorithm and the classified image is shown in Fig. 9(c). The objective function and parameters remain the same as before.

3. COMPARATIVE STUDY AND RECOMMENDATIONS

The overall results for all the images are given in Table 1. We recommend the distance transform method to the user industry. From a comparative study of the results given above, it follows that the fuzzy c-means classification approach is close behind the distance transform one, the gray-level morphological approach being least ideal for an user not much exposed to IP/PR techniques. The factor that matters the most is obviously the accuracy of the method. After extensive testing, a part of which are reported here, the distance transform method is found to give the most stable result and consistent with the human approach pre-



(a) (b)



(c)

Fig. 9. (a) The image in Fig. 4(a) after Hough transform and thresholding Hough space at 75 followed by calculation of local maxima. (b) The image in Fig. 6(a) after subtraction of the image in (a). (c) The image in (b) after applying fuzzy c-means clustering with $c = 3$, 2 features and fuzzy exponent $m = 2$.

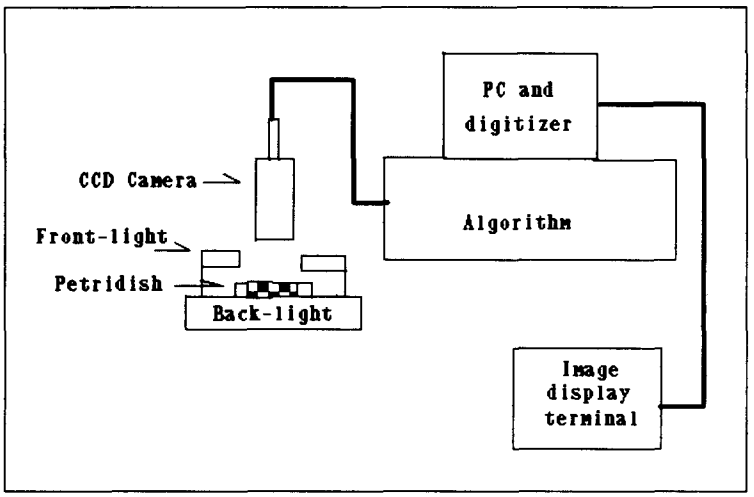


Fig. 10. Set-up for water quality analysis.

sently followed. The other key factors which cannot be ignored are the processing time per image and the degree of user interaction, particularly in selecting the various threshold values. In respect of processing time, the distance transform method is the fastest, followed by the fuzzy c-means and gray-level morphology approaches. However, if the Hough transform is to be added to detect the linear structure in the image, the performance of the fuzzy c-means classifier is expected to be worse compared to that with gray-level opening. The most ad hoc step is the threshold selection at the time of segmentation of the gray-level opened image

Table 2. Processing time including I/O on SUN3

Method	Processing time
Gray-level morphology	240 s
Distance transform	25 s
Fuzzy c-means clustering	204 s

as described in Subsection 2.3. The accuracy of threshold selection is very much operator-dependent and changes of selecting grid junctions as a potential colony is much higher, thereby reducing the performance of the system. On the other hand, thresholding of the distance-transformed image is quite straightforward and can be automated easily. In fact, the values of core points of grid lines, which we want to eliminate in distance-transformed space are constant. Though the processing time for the learning phase of the fuzzy c-means algorithm to calculate seed points of different clusters is quite high, the classification that follows then can be done instantly. However, we found that the use of seed-points of a set of images to an almost similar set is not always advisable and may give unsatisfactory results, thereby demanding learning even for a minor variation in the images. A list of processing times for comparison purposes is given below in Table 2. A set-up for water quality analysis is shown in Fig. 10.

4. FUTURE DIRECTION

We hope to achieve better performance by tuning several steps of the algorithms mentioned above. Some of the issues on which we are concentrating now are as follows:

(1) an improved and faster connected component labeling algorithm.

(2) Modified region growing algorithm as detailed in Subsection 2.4. Note that our objective here is to differentiate between a single colony grown to a larger size from a congruence of colonies.

(3) Tuning up of fuzzy c-means algorithm with a large number of samples. This process will be an attractive alternative if stabilized ready-to-use seed points for the colonies, the background and the grid lines could be provided.

(4) A fast linear structure detection algorithm using templates. Note that the orientation of the grid lines is not constant for all the samples due to the human factor involved in sample preparation and in placing the petridish under CCD camera. However, the grid lines will remain mutually parallel.

Acknowledgements—We would like to thank Dr Anjana Dewanji, Dr Nikhil Ranjan Pal, Dr Bhabatosh Chandra and Mr Punam Saha for their invaluable assistance in the course of this work, and Prof. B. B. Chaudhury for his constant encouragement. We would also like to thank Prof. D. P. Pal and Dr U. Chatterjee of the All India Institute of Hygiene and Public Health for providing samples and domain knowledge. Finally, we would like to record our appreciation of the cooperation extended by Dr R. N. Bhattacharya and Dr Bhunia of the Central Pollution Control Board, Calcutta.

REFERENCES

1. Health criteria and epidemiological studies related to coastal water pollution. Report of a group of experts jointly convened by WHO and UNEP. WHO Regional Office for Europe, Copenhagen (1977).
2. M. J. Pelczar, R. D. Reid and E. C. S. Chan, *Microbiology*, 4th edn. Tata McGraw-Hill, New Delhi (1983).
3. S. R. Sternberg, Grayscale morphology, *Comput. Vision, Graphics Image Process.* **35**, 333–355 (1986).
4. C. Arcelli and S. de Baja, A width-independent fast thinning algorithm, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-7**, 463–474 (1985).
5. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981).
6. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. Academic Press, New York (1982).
7. M. K. Chan, K. C. Kwok, Y. M. Shuen and K. L. Kool Development of a microcomputer-based system for the automatic recognition of *E. coli* colonies. *Laboratory Microcomputer* 95–101 (1990).

About the Author—DIPTI PRASAD MUKHERJEE has joined National Centre for Knowledge Based Computing at Indian Statistical Institute in 1991 after completing M.S. degree from the University of Saskatchewan, Canada. His responsibility includes the applications of image processing and computer vision in industrial inspection problems. He has recently submitted his Ph.D. thesis on detection and use of reflectional symmetry in computer vision. His research interest also includes development of expert system for industry.

About the Author—AMITA PAL (PATHAK) obtained her M.Sc. degree in statistics from the University of Calcutta in 1981, and her Ph.D. degree from the Indian Statistical Institute, Calcutta in 1991. She is working as a lecturer in the Computer Science Unit of the Indian Statistical Institute, Calcutta, and is currently visiting the Department of Mathematics of the Imperial College of Science, Technology and Medicine, London as a UNDP fellow. Her interests are mainly in the areas of pattern recognition and machine learning.

About the Author—S. ESWARA SARMA holds a degree in electronics and communication engineering. He has been involved in developing knowledge based systems for different applications for the last seven years. His research interests include the areas of image processing, pattern recognition and knowledge engineering.

About the Author—DWIJESH DUTTA MAJUMDER received the M.Sc.(Tech) from Calcutta University. In 1962 he received the Ph.D. degree in memory technology from the same university. In 1955, he joined the Electronic Computer Division of Indian Statistical Institute (ISI). From 1972 until 1992 he was head of the Electronics and Communication Sciences Unit in ISI. He has published more than 300 research papers and seven books in the fields of memory technology, speech and other types of pattern recognition, image analysis, computer vision, artificial intelligence, neural modeling, cybernetics and knowledge based computing. Dr Majumder is chairman of the National Centre for Knowledge Based Computing, Professor Emeritus of ISI, Distinguished HCL Professor of IIT, and Emeritus Scientist of CSIR. He is a fellow of INSA, INAE, IASc, IETE and is a member of the Governing Boards of International Association of Pattern Recognition, the World Organization of Cybernetics Systems and the International Fuzzy Systems Association.