Automated Analysis for Counting the Number of the Bacteria

Mu-Ya Horng[1], Shye-Chorng Kuo[2]

[1] Department of Environmental Engineering and Science, Chia Nan University of Pharmacy and Science
[2]Department of Multimedia Animation and Application, Nan Kai University of Technology

通訊作者：郭世崇
聯絡地址：542 南投縣草屯鎮中正路 568 號
電子郵件：t127@nkut.edu.tw
投稿日期：2018 年 4 月
接受日期：2018 年 6 月

Abstract

In recent years, to improve the environmental protection problems is vigorously developing in domestic and foreign areas and the pollution of water, soil, and air cause the environmental quality to deteriorate rapidly. Especially, the biological pollution that can not be inspected by eyesight and that could be the most dangerous and terrible problems within our environment. The colony of the bacteria is used to be the indicator of the environmental engineering water protection, and also to be the evaluation of the plan of reducing the environmental pollution, or the evaluation of the water quality and the sewage treatment. In the research of the colony of bacteria, counting the number of the bacteria needs observation and recognition by the experienced inspectors, and the tasks are very burdensome, though that can be implemented through the auxiliary counter. To decrease the squander on time and manpower, and also to reduce the erroneous recognition caused by manual counting, it is necessary to develop an objective tool to deal with the task and let the inspector finish the work quickly and accurately. In this paper, we exploit the image processing techniques and propose a pattern recognition method to complete the automated analysis for counting the number of bacteria system. These steps include image preprocessing, feature extraction, the recognition by extracted features, and the last is counting. By the results of the experiments, the average accuracy rate of our method in the automated analysis for counting the number of bacteria is 98%.

Keywords: colony of bacteria, image processing, pattern recognition

1. Introduction

Accurate counting the number of bacteria is crucial for quantitative analysis in water quality evaluation. This boring process by manual counting is difficult to implement for high-throughput assays due to low speed and inconsistent results.

The automation of counting the number of the bacteria has been interesting in recent years, and these methods have been shown to be more accurate than the manual counting (Barber, Vojnovic, Kelly, Mayes, Boulton, Woodcock, & Joiner, 2001; Herman, Pelgrim, Kirkels, Verheijen, Debruyne, Kenemans, & Vooijs, 1983; Lumley, Burgess, Billingham, McDonald, & Milligan, 1997) demonstrated that the automated counts had more accuracy than those manually determined by individual inspector. Similarly, (Parry, Chin, & Oonahoe, 1991) showed the significant difference in colony forming unit granulocyte-macrophage assay results which were taken from 4 laboratories in colony counting.

Some commercial products exist to promote accurate number counting, ranging from manual counting aids such as counting pens to all-in-one platforms including image acquisition, processing, and analysis. Nevertheless, most of the automated counting systems were capable of batch processing multiple images at once but the system were too complicated and time consuming. For these reasons, alternatives to commercial products using consumer-grade equipment, such as document scanners (Marotz, Lubbert, & Eisenbeiss, 2001; Biston, Corde, Camus, Marti-Battle, Esteve, & Balosso, 2003; Dahle, Kakar, Steen, & Kaalhus, 2004; Putman, Burton, & Nahm, 2005; Masala, Bottigli, Brunetti, Carpinelli, Diaz, Fiori, Golosio, Oliva, & Stegel, 2007; Bewes, Suchowerska, & McKenzie, 2008; Dobson, Reading, & Scutt, 1999) or digital cameras (Masala, Bottigli, Brunetti, Carpinelli, Diaz, Fiori, Golosio, Oliva, & Stegel, 2007; Wang, Yamaguchi, Someya, & Nasu, 2007; Sieuwerts, de Bok, Mols, de Vos, & Van Hylckama Vlieg, 2008; Chen & Zhang, 2009; Corkidi, Diaz-Uribe, Folch-Mallol, & Nieto-Sotelo, 1998) have been explored.

In this paper, the image processing technology is used to automatically count the number of bacteria. Apply the morphology technology to the binary image which is extracted from the image in the petri plate, and then eliminate noise to keep the important features. Based on the contour obtained from the binary image, we extract the shape features by the LSF (Least Square Fit) algorithm (Spath, 1997; Ahn, Rauh, & Warnecke, 2001; Gander, Golub, & Strebel, 1994). The shape feature is applied back to the original image and the area coverage is used as the screening method. By the area coverage rate, it can decide whether the contour is the bacteria or not and then complete the automatic counting.

The rest of the paper is organized as follows. Section 2 presents the method including image preprocessing, segmentation, edging, feature extraction, and counting. Section 3 shows the experimental results, and we conclude the paper in Section 4.

2. Method

According to the observations of the bacteria image, the shapes of the bacteria image approximate to ellipses which are close to a circle. In this study, we exploit the image processing techniques to deal with the pattern analysis for the bacteria in the petri plate, and then apply the Least Square Fit (LSF) algorithm to find a best curve to fit the outline of the bacteria pattern. We develop a method to extract the features of the bacteria in the image by the relationship between the fitting curve and the outline of the bacteria pattern, and then the extracted features are used as the key factors to recognize the bacteria. Finally, it executes the number of counting. The method is divided as following steps: (1) image preprocessing (2) image segmentation (3) feature extraction (4) recognition and the number of counting.

2.1 Image preprocessing

First, we acquire the bacteria image from petri plate by microscope, and the image type is gray scale with 200 dpi resolution and the gray scale image is with 8-bit gray level image containing gray level values from 0 to 255. Thresholding is a simple but effective method to distinguish objects from the background. A commonly used method, the Otsu method (Otsu, 1979), improves the image segmentation obviously. In Otsu method, the automatic binarization of gray level image that separates object from its background is accomplished depending on the threshold value chosen from the histogram of the image. Since it is effective and easier to implement, we apply the Otsu method to separate objects from the background. The original gray scale image is shown in Fig. 1(a) and the binary image through the Otsu method is shown in Fig. 1(b).

According to statistics of the bacteria image, the size of a bacteria image should larger than 30 pixels. Therefore, we apply the opening operation in morphology to remove all the objects containing fewer than 30 pixels that are regarded as noise, and the result is shown in Fig. 2.

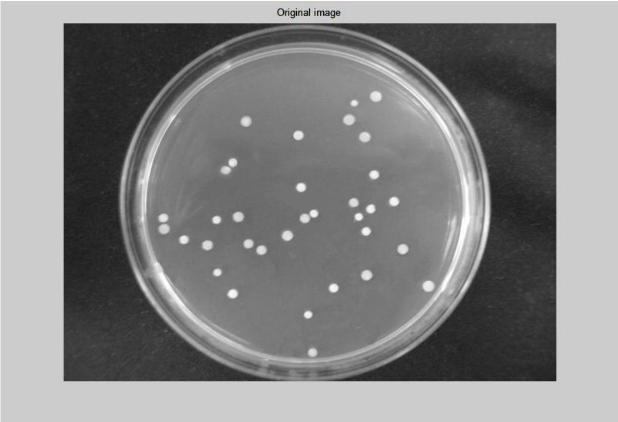

Fig. 1(a) Original image

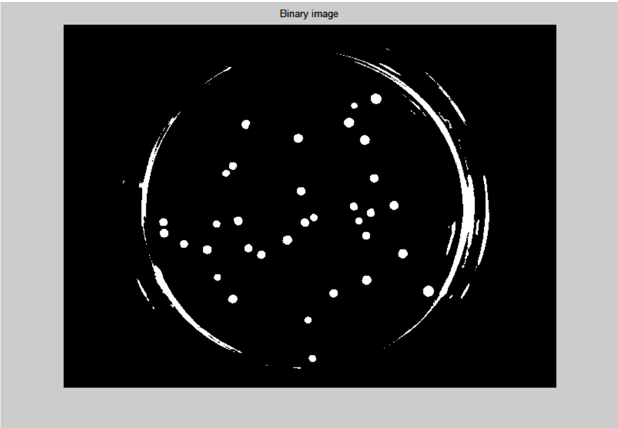

Fig. 1(b) Binary image through Otsu method

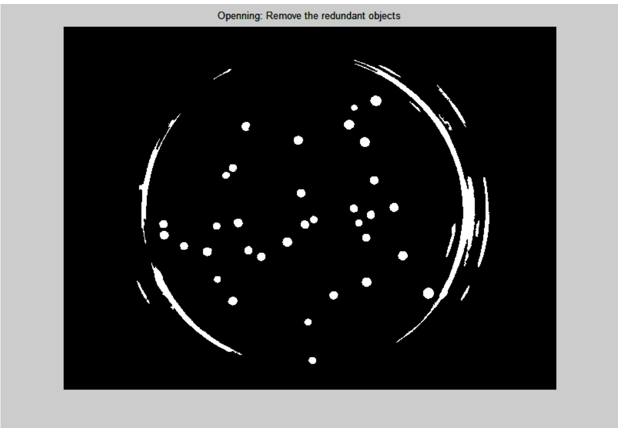

Fig. 2 Remove noise by opening operation

2.2 Segmentation

We apply the segmentation algorithm proposed by (Kuo S.C., Kuo S.S., Yu & Lee, 2007) to deal with the binary image segmentation. Since the algorithm can successfully segment the separated objects within the binary image, and the most important of all is that the algorithm can be easily implemented by the computer program languages because of the characteristics of operations on the stack, and is with short time consumption. After segmentation, the segmented objects are labeled with different colors, and the result is shown in Fig. 3.

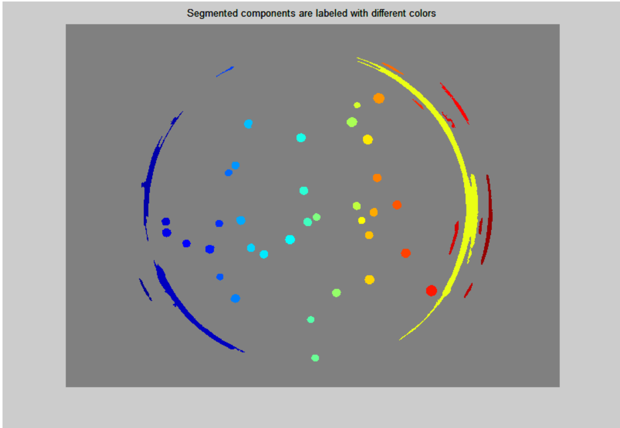

Fig. 3 Segmented objects labeled with different colors

2.3 Edging

Before performing the next step, a set of data points which constitute the edge of the object is extracted from the input gray scale image. After processing the segmentation, we convolve the images with the high-pass filter shown in equation (1) to extract the object's edge which forms a set of data points as the input data for the LSF algorithm.

$$HPF = \begin{bmatrix} -1/9 & -1/9 & -1/9 \\ -1/9 & 8/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \end{bmatrix} \tag{1}$$

We take one segmented object which is labeled with SC1 in the Fig. 3 as example as shown in Fig. 4, and it shows the binary image of SC1 in Fig. 5(a). The edge information of SC1 is shown in Fig. 5(b).

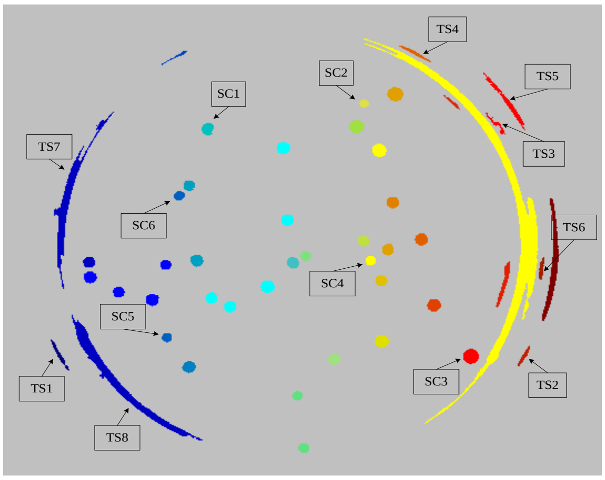

Fig. 4 Segmented objects with labels

Fig. 5(a) Binary image for SC1

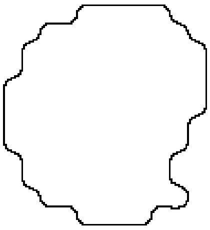

Fig. 5(b) Edge for SC1

2.4 Feature extraction

By the observation of the image patterns, the shapes of the bacteria are ellipses. Besides, those ellipses are close to the shape of a circle, though their sizes are not equal. Therefore, we utilize the edge information as the input data to feed the Least Squares Fitting (LSF) method to find an ellipse to fit the shape of the bacteria. Then, we consider two types of error to form the features for recognition. One is the coverage ratio, and the other is the mismatch ratio.

To begin with this algorithm, we assume the shape of the object is elliptic and an ellipse centered at the (x_c, y_c) can be expressed in parametric form as:

$$\begin{cases} x = x_c + a\cos(\alpha)\cos(t) - b\sin(\alpha)\sin(t) \\ y = y_c + a\sin(\alpha)\cos(t) + b\cos(\alpha)\sin(t) \end{cases} \quad (2)$$

,where a and b are the semi-lengths of the major and the minor axes respectively, α is the rotation angle of the ellipse, and t is the variable in the range between 0 and 2π.

There are many researches (Bookstein, 1979; Crawford, 1983; Chernov, & Ososkov, 1984; Karimäki, 1991; Thomas, & Chan, 1989) to develop ellipse detection by curve fitting. Least Squares Fitting (LSF) technique (Rosin, 1993) focuses on finding a set of parameters that minimize some distance measure between the data points and the ellipse. The implicit second order polynomial given in equation (3) is used to fit the data points to a conic and the constraint is used to represent the conic as an ellipse.

$$\varphi x^2 + \beta xy + \omega y^2 + dx + ey + f = 0, \quad \beta^2 - 4\varphi\omega < 0 \quad (3)$$

The direct least-squares technique (Halır & Flusser, 1998; Fitzgibbon, Pilu, & Fisher, 1999) is efficient for fitting ellipse in scattered data, and it can obtain the five parameters of the fitted ellipse, i.e., $x_c, y_c, a, b,$ and α. We apply this algorithm to find the fitted ellipse for SC1 as shown as Fig. 6. In this figure, the red dash curve represents the fitted ellipse constructed by LSF, and the black curve indicates the edge of SC1.

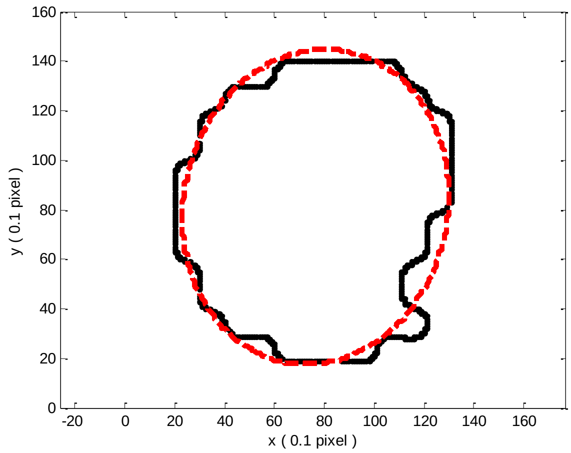

Fig. 6 The fitted ellipse for SC1

Next, we make a decision to recognize the shape of the object. First, we introduce two terminologies which are "covered area" and "non-match area" in this section. The covered area is defined as the number of the pixels that belong to the image object and are within the fitted ellipse. Non-match area due to object is defined as the number of the pixels that belong to the image object but are not within the fitted ellipse. Similarly, non-match area due to the fitted ellipse is defined as the number of the pixels that belong to fitted ellipse, but are not within the image object. Area of mismatch can be computed as the sum of non-match area due to image object and non-match area due to the fitted ellipse.

The foundation of this step is to define a coverage ratio that represents the portion which is bounded by the equation (4) in the object, and further to make a decision to determine whether the input object is an elliptical shape or not by this coverage ratio.

First, we set up a decision function which is derived from equation (2) as below,

$$f(x,y)=\frac{((x-x_c)\cos\alpha+(y-y_c)\sin\alpha)^2}{a^2}+\frac{(-(x-x_c)\sin\alpha+(y-y_c)\cos\alpha)^2}{b^2} \quad (4)$$

, where a and b denote the semi-major and semi-minor axis respectively, and α denotes the rotation angle.

For the coverage ratio, let Ω_k represent a set of pixels which are within the *k-th* image object after binarization and $N(\Omega_k)$ denotes the number of the elements in Ω_k. Let Ω_{kc} be a subset of Ω_k and define Ω_{kc} as follows:

$$\Omega_{kc}=\{(x,y)\mid f(x,y)\le 1,\ where\ (x,y)\in\Omega_k\} \quad (5)$$

Let λ_k be the coverage ratio defined for the *k-th* image object as follows:

$$\lambda_k=\frac{N(\Omega_{kc})}{N(\Omega_k)}\% \quad (6)$$

Next, for the mismatch ratio, let Ω_{km} defined as below:

$$\Omega_{km}=\{(x,y)\mid f(x,y)\le 1,(x,y)\notin\Omega_k\}\cup\{(x,y)\mid f(x,y)>1,(x,y)\in\Omega_k\} \quad (7)$$

Let ρ_k be the mismatch ratio that is defined for the *k-th* image object as follows:

$$\rho_k=\frac{N(\Omega_{km})}{N(\Omega_k)}\% \quad (8)$$

We use both the coverage ratio and the mismatch ratio to evaluate the recognition efficiency. The coverage ratio will be 100%, and the mismatch ratio will be 0% for a perfect circular or elliptical object for a good fitting. So the criterion is that the higher value of the coverage ratio and the lower value of the mismatch ratio are better approach for the recognition.

Take the SC1 for example. In Fig. 7, it shows the result for the coverage area which is labeled the color blue, and the coverage ratio λ_k is 95.67%. In Fig. 8, it shows the result for the non-match area which is labeled the color green, and the mismatch ratio ρ_k is 12.21%.

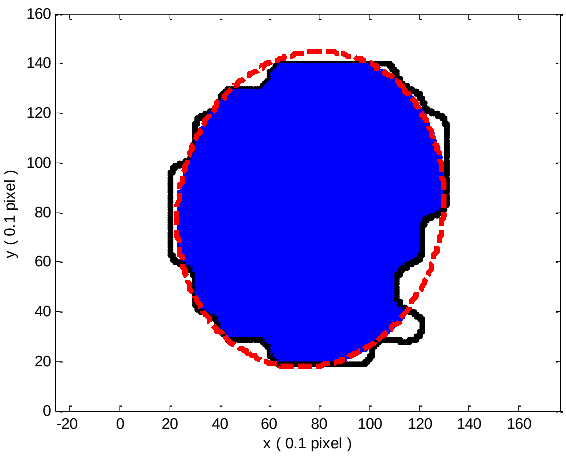

Fig. 7 The coverage area for SC1

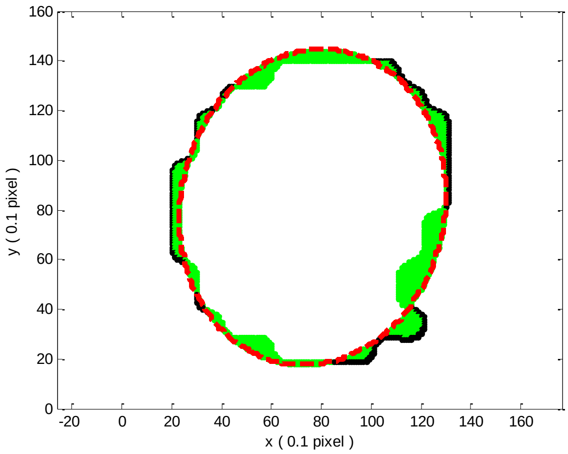

Fig. 8 The non-match area for SC1

In addition to the coverage ratio and the mismatch ratio, we take the ratio of semi-minor axis to semi-major axis as another feature, and we call it the flat ratio of the object. If the flat ratio is more close to 1, the shape of the object is more close to a circle. In the example SC1, we obtain 5.301 (pixels) and 6.401 (pixels) for the semi-minor axis and semi-major axis respectively, and the flat ratio is 82.82%.

Then, take the TS6 for another example and the fitted ellipse for TS6 is as shown as Fig. 9. In Fig. 10, it shows the result for the coverage area which is labeled the color blue, and the coverage ratio λ_k is 92.64%. Meanwhile, In Fig. 11, it shows the result for the non-match area which is labeled the color green, and the mismatch ratio ρ_k is 26.49%. Next step, we take the flat ratio of TS6 and the value is 15.58% which is too low to be a circle and is identified not to be a bacteria.

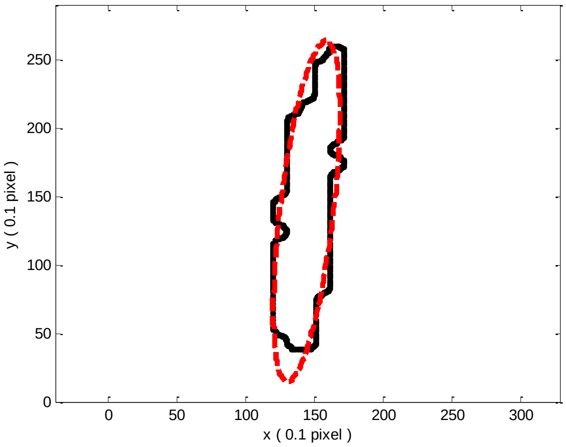

Fig. 9 The fitted ellipse for TS6

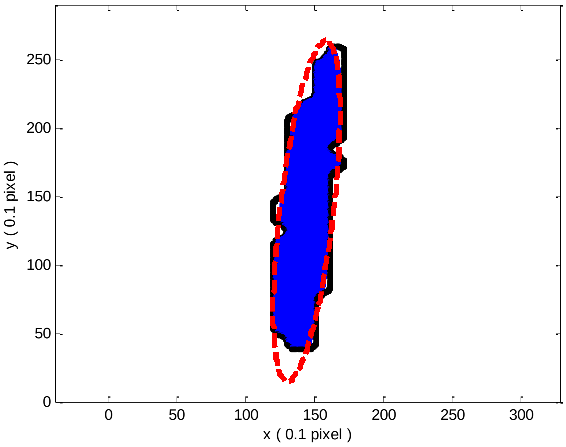

Fig. 10 The coverage area for TS6

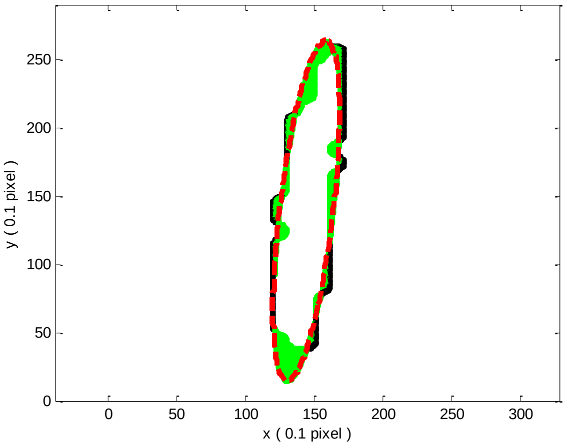

Fig. 11 The non-match area for TS6

3. Experimental Results

To recognize the bacteria is to use the three features: coverage ratio, mismatch ratio, and flat ratio. By our experiments, the criterion is as following, if the coverage ratio is larger than 90%, the mismatch ratio is less than 15%, and in addition the flat ratio is larger than 80%, then the object is recognized as the bacteria.

In this case of the image in Fig. 1, we obtain the number of bacteria in the image is 34, and add the red cross mark to the objects which are recognized as the bacteria in the image as shown in Fig. 11. The accurate rate is 100%.

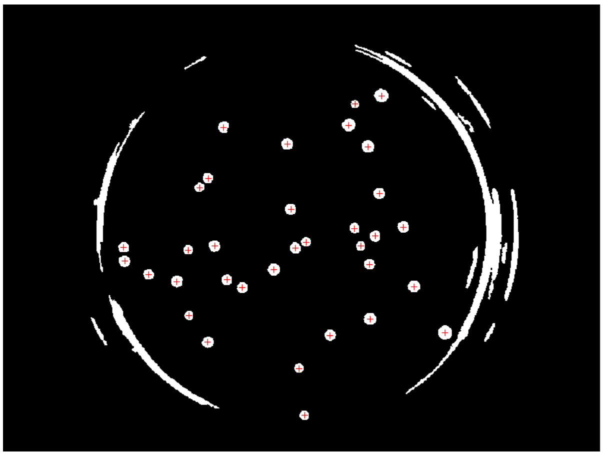

Fig. 11 The found bacteria in the image

Table 1 The features of the several objects labeled in the image in Fig. 1

label of the object	coverage ratio(%)	mismatch ratio(%)	a (pixels)	b (pixels)	b/a (%)
SC1	95.67	12.21	6.40	5.30	82.82
SC2	96.74	9.97	4.34	4.07	93.71
SC3	98.41	4.92	7.39	7.13	96.46
SC4	98.05	7.45	4.90	4.69	95.68
SC5	96.63	10.76	4.67	4.59	98.31
SC6	98.05	6.68	5.18	4.47	86.27
TS1	92.09	28.81	18.81	1.70	9.01
TS2	94.58	20.82	12.06	1.79	14.88
TS3	77.13	103.28	16.35	1.91	11.68
TS4	93.16	27.62	17.76	1.58	8.87
TS5	87.86	61.88	67.69	2.12	3.13
TS6	92.64	26.49	12.53	1.95	15.58
TS7	54.88	2065.80	111.56	54.42	48.78
TS8	59.01	732.85	86.54	30.92	35.73

In Table 1, we list the three features of several objects labeled in the image in Fig. 1. The labels which are named as SC1~SC6 represent the bacteria, and the others (TS1~TS8) indicate the redundant objects. We also test another 25 images of bacteria which are acquired from the microscope on different time. All the testing images are gray scale with 200 dpi resolution, and the gray scale image is with 8-bit gray level image containing gray level values from 0 to 255. The proposed method is implemented by using Python on Intel® Core™ i5, 3.3GHz with 8G RAM. The average accuracy rate is 98%. The primary errors are caused by the flaws of the image, such as the focus of the photography is not accurate, and the difference of the gray scale between the bacteria and background is too small in image. These two factors should affect the results in segmentation process and distort the shape of the bacteria, and then it causes the errors.

4. Conclusion

At present, it needs observation and recognition by the experienced inspectors or researchers to count the number of the bacteria. They often use their eyesight to identify the bacteria in the slide and to count the number of bacteria by auxiliary counter. Since the bacteria scatter in the slide, it is a hard work to do the task exactly. We proposed a method to resolve this problem. The executing time of the process was short since there were no complicated and iterative instructions in the algorithm. We developed this system that could decrease the squander on time and manpower, and also decrease the erroneous recognition caused by manual counting.

In our experiments, we took off the images which contained the overlap bacteria, i.e., the bacteria over the edge of another. Therefore, In the future work, we will study how to solve this problem, and we will apply this method to recognize the objects with elliptical shape.

References

Ahn S.J., Rauh W., & Warnecke H. (2001). Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. *Pattern Recognition*, 34(12), 2283-2303.

Bookstein F.L. (1979). Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9(1), 56–71.

Barber P.R., Vojnovic B., Kelly J., Mayes C.R., Boulton P., Woodcock M., & Joiner M.C. (2001). Automated counting of mammalian cell colonies. *Physics in Medicine & Biology*, 46(1), 63-76.

Biston M.C., Corde S., Camus E., Marti-Battle R., Esteve F., & Balosso J. (2003). An objective method to measure cell survival by computer-assisted image processing of numeric images of Petri dishes. *Physics in Medicine & Biology*, 48(11), 1551-1563.

Bewes J.M., Suchowerska N., & McKenzie D.R. (2008). Automated cell colony counting and analysis using the circular hough image transform algorithm (CHiTA). *Phys Med Bio*, 53, 5991-6008.

Crawford J.F. (1983). A non-iterative method for fitting circular arcs to measured points. *Nuclear Instruments and Methods in Physics Research*, 211(1), 223–225.

Chernov N.I., & Ososkov G.A. (1984). Effective algorithms for circle fitting. *Comp. Phys. Comm.*, 33, 329–333.

Corkidi G., Diaz-Uribe R., Folch-Mallol J.L., & Nieto-Sotelo J. (1998). Covasiam: An image analysis method that allows detection of confluent microbial colonies and colonies of various sizes for automated counting. *Applied and Environmental Microbiology*, 64, 1400-1404.

Chen W.B., & Zhang C.C. (2009). An automated bacterial colony counting and classification system. *Information Systems Frontiers*, 11(4), 349-368.

Dobson K., Reading L., & Scutt A. (1999). A cost-effective method for the automatic quantitative analysis of fibroblastic colony forming units. *Calcified Tissue International*, 65(2), 166-172.

Dahle J., Kakar M., Steen H.B., & Kaalhus O. (2004). Automated counting of mammalian cell colonies by means of a flat bed scanner and image processing. *Cytometry Part A*, 60A, 182-188.

Fitzgibbon A., Pilu M., & Fisher R.B. (1999), Direct least square fitting of ellipses, *IEEE Trans. PAMI*, 21(5), 477–480.

Gander W., Golub G.H., & Strebel R. (1994). Least-squares fitting of circles and ellipses. *BIT*, 34(4), 558-578.

Herman C.J., Pelgrim O.E., Kirkels W.J., Verheijen R., Debruyne F.M.J., Kenemans P., & Vooijs G.P. (1983). In-use evaluation of the omnicon automated tumor colony counter. *Cytometry*, 3, 439-442.

Halır R. & Flusser J. (1998). Numerically stable direct least squares fitting of ellipse. *In Proceedings of the 6th Conference in Central Europe on Computer Graphics and Visualization*, 253–257.

Karimäki V. (1991). Effective circle fitting for particle trajectories. *Nuclear Instruments and Methods in Physics Research*, 305(1), 187–191.

Kuo S.C., Kuo S.S, Yu Y.H. & Lee M.T. (2007). A Binary Image Segmentation Algorithm Using Dynamic Stack. *Journal of Nan-Kai*, 4(3), 9-14.

Lumley M.A., Burgess R., Billingham L.J., McDonald D.F., & Milligan D.W. (1997). Colony counting is a major source of variation in CFU-GM results between centres. *British Journal of Haematology*, 97, 481-4 84.

Marotz J., Lubbert C., & Eisenbeiss W. (2001). Effective object recognition for automated counting of colonies in Petri dishes. *Computer Methods and Programs in Biomedicine*, 66, 183-198.

Masala G.L., Bottigli U., Brunetti A., Carpinelli M., Diaz N., Fiori P.L., Golosio B., Oliva P., & Stegel G. (2007). Automatic cell colony counting by region-growing approach. *Nuovo Cimento*, 30, 633-644.

Otsu N. (1979). A Threshold Selection Method from Gray-Level Histogram. *IEEE Trans. Systems Man, and Cybernetics*, 9, 62-66.

Parry R.L., Chin T.W., & Oonahoe P.K. (1991). Computer-aided cell colony counting. *Biotechniques*, 10, 772-774.

Putman M., Burton R., & Nahm M.H. (2005). Simplified method to automatically count bacterial colony forming unit. *Journal of Immunological Methods*, 302, 99-102.

Rosin P.L. (1993). A Note on the Least Squares Fitting of Ellipses. *Pattern Recognition Letters*, 10(14), 799-808.

Spath H. (1997). Least squares fitting of ellipses and hyperbolas. *Computational Statistics*, 12, 329-341.

Sieuwerts S., de Bok F.A.M., Mols E., de Vos W.M., & Van Hylckama Vlieg J.E.T. (2008). A simple and fast method for determining colony forming units. *Letters in Applied Microbiology*, 47, 275-278.

Thomas S.M., & Chan Y.T. (1989). A simple approach for the estimation of circular arc center and its radius. *Computer Vision, Graphics and Image Processing*, 45(3), 362–370.

Wang X.D., Yamaguchi N., Someya T., & Nasu M. (2007). Rapid and automated enumeration of viable bacteria in compost using a micro-colony auto counting system. *Journal of Microbiological Methods*, 71, 1-6.

生菌數量之自動計數

洪睦雅[1]、郭世崇[2]

[1]嘉南藥理大學　環境工程與科學系
[2]南開科技大學　多媒體動畫應用系

摘　　要

近年來國內外環保問題日趨嚴重，水質、土壤、空氣之污染造成環境品質急遽惡化。其中生物性污染非人類肉眼可見，可謂是最危險及最可怕的環境污染問題。生菌菌落常作為環境工程水質污染程度之指標微生物，以及水、廢污水處理效率之評估。在從事總生菌數研究時，計數數量雖可藉由菌落計數器計數，但仍需要人力觀察辨識與做計數的繁雜工作，為了降低在時間及人力上的大量耗費，並減少因人為的因素所產生的誤判，因此極需客觀的資訊工具來處理這些過程，讓檢驗者以最迅速的方式來完成辨識與正確的計數工作。本研究運用影像處理技術及提出一個圖形辨識的方法來完成生菌數量之自動計數系統，主要步驟為影像前處理、擷取特徵值與利用所擷取的特徵值做辨識，最後做自動計數。由實驗結果所得之生菌數量之自動計數的正確率為 98%。

關鍵詞：生菌菌落、影像處理、圖形辨識