A Novel Bacteria Classification Scheme Based On Microscopic Image Analysis

LI XIAOJUAN

Information Engineering College

Capital Normal University

Beijing, China

CHEN CUNSHE

S.C.E.E.

B.T.B.U.

Beijing, China

YV ZIYI

Information Engineering College

Capital Normal University

Beijing, China

Abstract: - A novel image of bacteria classification scheme is presented. ITSMM is proposed for the edge detection of bacteria images, which effectively remove or restrain noise and get clear edges of microscopic image by optimizing segmentation threshold and changing the order of edge detection. Principal Component Analysis (PCA) is used to reduce the dimensionality of original features extracted from the images and BP network is used to learn to classify the images. The proposed method is tested with NMCR databases, experimental results demonstrate the presented image classification scheme is effective.

Key-Words: - bacteria Image Analysis , edge detection, PCA

1 Introduction

A customary procedure used to determine bacterial abundance, biovolume and morphology is microscopic examination of filtered fluorescently dyed cells[5]. Identification, counting and measuring individual cells based on exampling are normally done manually, which is tedious and time-consuming. However, image analysis is a efficient way to do that. In the paper a novel image of bacteria classification scheme is presented, which can be used for versatile, efficient processing of many images in order to obtain reliable, high-resolution data of microorganisms. ITSMM is proposed which is effective for the edge detection of bacteria images by optimizing threshold value for segmentation combining with mathematicl morphology method, which effectively remove or restrain noise and get clear edges of microscopic image. Principal component analysis (PCA) is a useful method for multivariate analysis. In the paper it is used to reduce the dimensionality of original features extracted from the images and a improved BP network is introduced to train for the bacteria images classification. A effective bacteria image classification can greatly improve the speed and consistency in performing large-scale surveys or rapid determination of bacterial abundance, morphology and makes sampling at a higher resolution than is practical manually.

2 ITSMM Edge Detection for Bacterial Image

. The edge is one of the fundamental features for digital images, which is the basis of image segmentation, features extraction and recognition for aim area. The edges of objects become blurred when they are viewed at high magnification, which is caused by the limited resolution of microscope optics. Moreover, uneven lighting and differences in exposure times and object luminescence make it impossible to choose a single gray level as a threshold for distinguishing objects from the background. Viles and Sieracki (3) discussed this problem with respect to bacteria under epifluorescence conditions and concluded that a Marr-Hildreth operator functions with a high degree of independence for exposure and lighting characteristics and with accurate edge detection properties. A Marr-Hildreth operator is a combination of a Gaussian operator for smoothing and a Laplace operator for amplifying high spatial frequencies. The Laplace operator calculates the second derivative of intensity. If the edge of a blurred object is at the point where the rate of intensity change is greatest (second derivative), the Laplace operator identifies its position as the zero crossing between positive and negative values. However, the Laplace operator is sensitive to electronic camera noise and to faint particles, which can be a problem for analysis of the images with noisy singals . So we propose ITSMM(iterative threshold segmentation and mathematical morphology), a method of edge detection for bacteria images, which is combined the iterative threshold value segmentation with mathematicl morphology edge detection.

2.1 Iterative Threshold Value Determination

The iterative improved policy on threshold valuei s the key factor of the algorithem. Initially a approximation threshold value is selected, by which a image is segmented and produces sub-images, and then a new threshold value is selected dependent on the features of the sub-images. The new threshold value is continues used to segment the images and so on, until the error pixel of the images is minimized:

(1). Initially a estimated threshold value is determined:

$$T^0 = \left\{ T^K \mid k = 0 \right\}$$

$$T_0 = \frac{Z_{min} + Z_{max}}{2}$$

(6)

Z_{mn} Z_{max} is respectively the minimum and maximum gray level of the image。

（2）.The Image is divided into two areas(R1 and R2)by T^k

$$R1 = \left\{ f(x,y) \mid f(x,y) \geq T^k \right\}$$

$$R2 = \left\{ f(x,y) \mid 0 < f(x,y) < T^k \right\}$$

(7)

(3). figure out the median of gray level of area R1 and area R2

$$Z_1 = \frac{\sum_{f(i,j)<T^k} f(i,j) x N(i,j)}{\sum_{f(i,j)<T^k} N(i,j)}$$

(8)

$$Z2 = \frac{\sum_{f(i,j)>T^k} f(i,j) x N(i,j)}{\sum_{f(i,j)>T^k} N(i,j)}$$

(9)

f(j,j) is the gray level at pixel（i,j），N（i,j）is weigth factor

（4）.A new threshold value T^{k+1} is determined

$$T^{k+1} = \frac{Z_1 + Z_2}{2}$$

(10)

if $T^k = T^{k+1}$,the segmentation is completed，else K ＝K+1,go to (2)。

Edge detection for identification of objects was performed with the optimal threshold value, which produced a binary image. Most of the background noise was effectively removed.

2.2 Iterative Threshold Value Determination

Although the process of threshold value segmentation remove or restrain most of the background noise by the optimal threshold value, it might be falsely associated with the edges of objects, and result in the blurred edge of image. Mathematics

morphologic methods is a useful for pre-processing or image analysis based on the flat structuring element operation. the structuring element operator is vital to result of the image process.in the paper the erosion operation in mathematical morphoplogy is used to improve the edge clarity of images, and erosion operation is performed on the binary image with 3x3 structure element, by which image edges with only one- pixel width is attained and results in the binary image used for measurement and classification, as shown in figure 4.

Analysis and a series of tests(NMCR database) show that the edge detection of bacteria microscopic image based on iterative threshold value segmentation and mathematical morphology can not only effectively remove or restrain noise but also get clear edges of microscopic image by determining proper threshold value for segmentation and changing the order of edge detection. Figure1 shows a original microscopic image of bacteria, figure 2 is the result of edge detection with Sobel operator in which the noise is blended among edge pixel and might result in unstable edge. Figure 3 is the result of edge detection with Gauss-Laplace, some of edge information about aim object is lost and made the edge is noncontiguous or isolated for that Laplace operaor is sensitive to electronic camera noise and to faint particles.
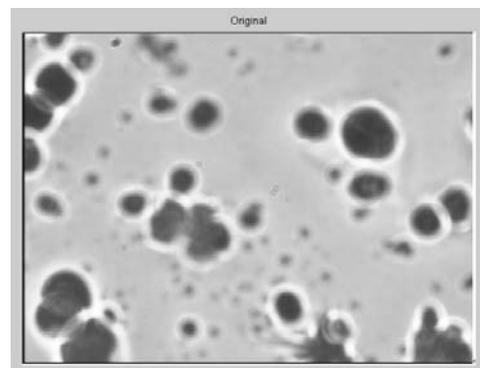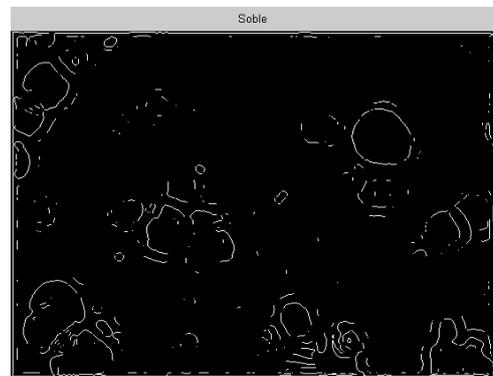
figure 1 the original image

figure2 edge detection with soble operator
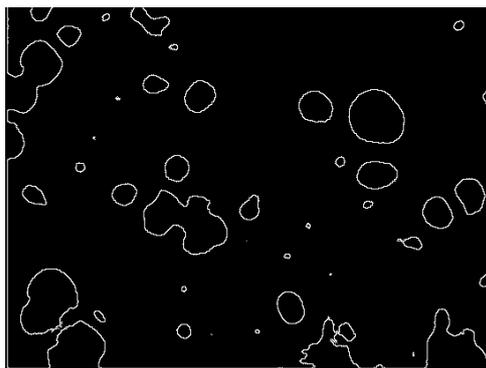
figure3 Gauss-Laplace edge detection

figure 4 edge detection with ITSMM

Edge detection we proposed for bacteria identification intensify the aim object of microscopic image and enhance the clarity of edges by iterating for proper threshold value while exactly locate the aim objects and segment it from background that achieved by combining the iterative threshold value segmentation theory with mathematicl morphology edge detection. The method of edge detection is noise-proof robust, and can be well balanced between refraining noise and the retained whole fine edges of aim objects.

3 Extraction Features of Microscopic Image

The original features which effectively describe the information of aim bacteria have to be abstracted from binary image after segmentation. There are many kinds of features for   construction or recognition a object by it's image, i. e. geometry features, algebra features, gray level statistics features, texture features and transformation space features etc. Research on a amount of bacteria images shows that there is great difference in shape and morphology feature among different bacteria, moreover in bacteria image analysis a bacteria is represented by twenty features, that is projection area, parimere, major axis,

minor axis, complexity, rectangular, seven invariant moments, which blong to shape and morphology features, and histogram mean, standard deviation, contrast of gray level , moment of the second order, uniformity, differential moment, entropy, which belongs to texture features. Because of such a high dimensionality of the image space and the limited dataset sizes, in the paper Principal Component Analysis (PCA) is used totransform from the image space onto a low dimensional space with the hope that a decision boundary for classification is easier to find in the low dimensional space.

3.1 Principal Component Analysis (PCA)

PCA is a standard technique in data analysis which is used for dimensionality reduction or equivalently for feature extraction for signal representation. Given a set of l data vectors $x_i \in R^n$ which are instances of random vector xi, PCA looks for m<n orthonormal vectors $\{\theta_j\} \in R^n$ which form an orthonormal basis of the subspace that captures maximal varianceof the xi's. It can be shown [37] that the $\{\theta_j\}'s$ are the eigenvectorsof the sample covariance matrix $\sum$ of the xi's

$$\sum = \frac{1}{l}\sum_{i=1}^{L}(x_i - \bar{x})(x_i - \bar{x})^T$$

Where $\bar{x}$ is the sample mean of the xi's. The approximation or reconstruction $\hat{x}$ of the random vector x is the component of x that lies in the subspace expressed in R^n i.e.

$$\hat{x} = \bar{x} + \sum_{j=1}^{m}(y_i - \bar{y}_j)\theta_j$$

Where $y_i = \theta_i^T x$ Is called the ith principal component of x and $\bar{y}_i$ is the ith principal component of the mean $\bar{x}$,i.e.

$$\bar{y}_i = \theta_i^T \bar{x}$$

By defining the n×m matrix Φ whose columns are m eigenvectors θ_j ofΣ we can write the principal component vector y ∈R^m

$$y = \theta^T x$$

and because the columns of θ are orthonormal, we have

$$\hat{x} = \bar{x} + \theta(y - \bar{y})$$

where $\bar{y} = \theta^T \bar{x}$. A very nice property of PCA is

that the mean square error ε^2 between x and its reconstruction, which is

$$\varepsilon^2 = E[\| x - \hat{x} \|^2],$$

Can be written as

$$\varepsilon^2 = \sum_{j=m+1}^{n} \bar{\lambda}_j$$

Where $\bar{\lambda}_j$ are the eigenvalues of the true(unknown) covariance matrix generating the x_i's. This last equation suggests that the mean square error between

x_i and its reconstruction $\hat{\lambda}_i$ is minimised if the subspace basis contains the m eigenvectors θ_j with the m highest eigenvalues.

The PCA decorrelates the x_i's as the covariance expressed in the principal subspace

$$\Sigma y = \theta^{T} \Sigma \theta$$

Is a m×m diagonal matrix whose diagonal elements are the λ_j's.

3.2 **Determining the features of the principal subspace**

Principal component analysis (PCA) is a useful method for multivariate analysis. The dimensionality 20 of the original bacteria can be reduced to p using PCA. On knowing the eigenvectors, one can transform the vectors into the eigenbasis. The components of t he new vectors are the projections of the old ones onto the eigenvectors. By algebra or a geometry criterion for optimization, We try to describe and to simplify a data matrix of image features, which is represented by some *m* dimensional vectors. If the *p* most relevant eigenvectors are enough to describe the features of a type of bacteria image, the vectors after the *p*th component is discarded. Thus the data with *p* dimensional for the features of a type of bacteria image can be attained. The choice of the dimensionality p of the principal is based on the characteristics of the eigenvectors so as to minimize the classification error rate on a data set. It is found that the intrinsic dimensionality of the type bacteria space depends on the preprocessing of the images. Using PCA, 8 features of bacteria imageis picked out of the 20 original features by Bayesian model selection for classification decision in the paper :
   a) area: the area of subset S is the total number of pixel in the S.
   b) parimer: the contour length of a object aim which is caculated by contour trace alogrithems of *image.*
   c) *major axis: the longest dimension of the aim object*
   d) *minor axis*: the Breadth the aim object
   e) *contrast:* $\sum_i \sum_j (i - j)^2 P_\delta(i, j)$, the mean of pixel pairs which gray level is $(|i - j|)$, $P_\delta(i, j)$ is the gray level at (i,j)
   f) *moment of the second order(energy):*

$$\sum_i \sum_j \{P_\delta(i, j)\}^2$$

   g) *Inverse differential moment:*

$$\sum_i \sum_j \frac{P_\delta(i, j)}{1 + (i - j)^2}$$

   h) *Entropy:* $-\sum_i \sum_j P_\delta(i, j) \lg P_\delta(i, j)$

Table 1 and table 2 below show the features description of 4 types of bacterial image respectively:
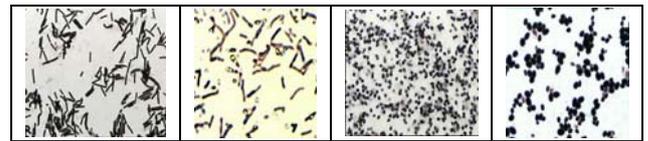

figure 5 four types of bacterial image Bacteria from NMCR database

TABLE I
MORPHOLOGICAL FEATURE S OF 4
TYPES BACTERAIA   IMAGES

	Bacteria 1	Bacteria 2	Bacteria 3	Bacteria 4
area	603	740	318	204
parimer	64.234	41.347	30.476	26.841
major axis	94.764	84.921	26.572	25.742
minor axi	19.562	7.5.606	22..650	31.769

TABLE 2
TEXTURES FEATURE S OF 4 TYPES BACTERAIA
IMAGES

	Bacteria 1	Bacteria 2	Bacteria 3	Bacteria 4
CGL	0.25617	0.2512	0.52562	0.09894
energy	0.08974	0.12531	0.12921	0.18983
Entropy	0.94216	0.92836	0.69621	0.91312

3.3 **Objects Classification Based on Improved BP Neural Network.**

Animal vision can elegantly accomplish classification of objects to a remarkable degree. Artificial can be used to mimic this process in machine vision, and the ability of an artificial network to learn can be measured by its ability to generalize beyond the examples that it is given during the training procedure[3]. The improved algorithm settles two problems in training a BP network : (1) introduce momentum factor in order to accelerate convergence rate. In training traditional

BP network learning rate η is fixed value which could cause slow convergence rate, we can increase η to accelerate convergence rate, but lead to unstable and if we avoid oscillation by decreasing η, it result in slow convergence rate. Therefore momentum factor is introduce $\Delta W(t+1) = \Delta W(t+1) + \alpha \Delta W(t)$ α is momentum coefficient $0 < \alpha < 1$。

(2) smooth drive function. weights modification of both hidden layer and output layer is related to y_s, the output of mid-layer; when y_s is near to 0 or 1 the weights modification approach 0 and the network get into local mimnum. In order to accelerate convergence rate, let y_s =0.1   when y_s <0.1, let y_s =0.9 when y_s >0.9.

a three-layer feedforward network layer is used in the paper. The eight features determined above extracted respectively from the eight types of bacteria is used to classify the types of bacteria. 60 samples are selected from NMCR database for training set, and 20 images from the database is used to test. The result shows that the accuracy of recognition rate reach 82.3%.

4 Conclusion

It is presented that a novel image of bacteria classification scheme which is proved effective for identification or determination of bacterial abundance and morphology. ITSMM is proposed for the edge detection of bacteria images, which intensify the aim object of microscopic image and enhance the clarity of edges . It is performed by optimizing for proper threshold value while exactly locate the aim objects and segment it from background which is achieved by combining the iterative threshold value segmentation theory with mathmaticl morphology edge detection. The method of edge detection is noise-proof robust, and can be well balanced between refraining noise and the retained whole fine edges of aim objects. Principal Component Analysis (PCA) is used to reduce the dimensionality of original features extracted from the images and BP network is improved in order to train adaptable for the bacteria images classification. Experimental results show that the presented image classification scheme is effective, The proposed method is tested with NMCR databases, experimental results demonstrate the presented image classification scheme is effective, which can greatly improve the speed and consistency in performing large-scale surveys or rapid determination of bacterial abundance, morphology and makes estimation of bacterial condition more accuracy.

References:
1) Guop ing Qiu, Xia Feng, J ianzhong Fang, Comp ressing Histogram Representations for Automatic Colour Photo Categorization, Pattern Recognition 37 : 2177 ∼2193, 2004
2) Sungyoung Kim, Sojung Park, and Minhwan Kim, Image Classification into Object /Non object Classes, CIVR 2004, LNCS 3115, 393 ∼ 400, 2004.
3) Massana, R., J. M. Gasol, P. K. Bjørnsen, N. Blackburn, Å. Hagstro ¨m, S. Heitanen, B. H. Hygum, J. Kuparinen, and C. Pedro´s-Alio . Measurements of bacterial size via image analysis of epifluorescence preparations: description of an inexpensive system and solutions to some of the most common problems. Sci. Mar. 61:397–407. 1997.
4) Zhang Yujin   Image Engineering(II) Image Analysis Tsinghua University Press , 2005.
5) J.F. Drouin, L. Louvel, B. Vanhoutte, H. Vivier, M.N. Pons, P. Germain,Quantitative characterization of cellular differentiation of Streptomyces ambofaciens in submerged culture by image analysis, Biotechnology Techniques, 11, 819-824 ,1997.
6) Cao L J. Chua K S, Chong W K, et al. A comparison of PCA , KPCA , and ICA for dimensionality reduction in support vector mach ine [J]. N eurocomput ing, 55 (1- 2) : 321 - 336. 2003.