Automatic Recognition of Mycobacterium Tuberculosis Based on Active Shape Model

Chao XU†, Student Member, Dongxiang ZHOU††, Tao GUAN†, Yongping ZHAI†††, and Yunhui LIU††††, Nonmembers

SUMMARY  This paper realized the automatic recognition of Mycobacterium tuberculosis in Ziehl-Neelsen stained images by the conventional light microscopy, which can be used in the computer-aided diagnosis of the tuberculosis. We proposed a novel recognition method based on active shape model. First, the candidate bacillus objects are segmented by a method of marker-based watershed transform. Next, a point distribution model of the object shape is proposed to label the landmarks on the object automatically. Then the active shape model is performed after aligning the training set with a weight matrix. The deformation regulation of the object shape is discovered and successfully applied in recognition without using geometric and other commonly used features. During this process, a width consistency constraint is combined with the shape parameter to improve the accuracy of the recognition. Experimental results demonstrate that the proposed method yields high accuracy in the images with different background colors. The recognition accuracy in object level and image level are 92.37% and 97.91% respectively.

key words: object recognition, tuberculosis, active shape model, watershed, segmentation

1. Introduction

9 million new cases of tuberculosis (TB), caused by Mycobacterium tuberculosis (MTB), were discovered in 2013 causing 1.5 million deaths[1]. TB has been the second killer worldwide only to HIV/AIDS due to a single infectious agent. More than 95% of TB deaths occur in low-and middle-income countries due to poor hygienic conditions and limited medical preventive measures. Early discovery and treatment is important for the control and treatment of TB. Currently, the most common and accurate method for diagnosing TB is manual screening, by which the trained laboratory technicians must examine a sputum smear slide under a microscope at 300 views. This approach is labor-intensive and leads to a high false negative rate when the laboratory technicians have to examine a large number of samples a day. With the rapid development of image processing and pattern recognition, computer-aided diagnosis for TB appeared. The images can be captured from sputum smear slides by microscopy autofocusing. Then the slides will be diagnosed by analyzing the captured images with related techniques of image processing and pattern recognition.

In 1998, Veropoulos et al. [2] first used computer technology to detect MTB objects. They studied the MTB images captured by a CCD camera mounted on a fluorescence microscope. Their approach first detected the edges of the objects in the gray images using Canny operator. Then the object regions were extracted by region labeling and removal, edge pixel linking and boundary tracing. 15 Fourier descriptors were used to represent the extracted region. Finally, the objects were recognized by a back-propagation (BP) neural network classifier. Their approach yielded a sensitivity of 93.53% and the specificity of 98.79%.

In 2001, Forero et al. [3] adopted multi-thresholding fuzzy segmentation in the color space, and this approach was later improved [4], [5]. In 2004, this group [6] adopted adaptive color thresholding in green channel, and combined the segmented results with the edges detected by Canny operator in the gray channel. The first four Hu moments and compactness were selected as the feature descriptors. The recognition was carried out by the k-means clustering and decision tree. Based on the former research, they [7] employed a Gaussian Mixture Model to model the results of k-means clustering and adopted a Bayesian classifier to classify the target objects. In 2012, Chang et al. [8] started to step into this area based on digital microscopes, CellScope. They applied a white top-hat transform and template matching of a Gaussian kernel to extract candidate objects. Each candidate was characterized by 102-dimensional feature vector including Hu moments, geometric and photometric features, and histograms of oriented gradients. Then the recognition was performed by a support vector machine classifier with an average accuracy of 89.2% ± 2.1%.

The sputum smear slides are stained with auramine-O in aforementioned work, and the images are captured by the fluorescence microscope. Although the fluorescence microscope is sensitive for MTB objects detection, it is high-cost with difficult equipment maintenance for developing countries. Thus other groups began to research on the images captured from Ziehl-Neelsen (ZN) stained slides with a conventional light microscope [9], [10]. Khutlang et al. [11] had made great progress in this area. They used a com-
bination of several pixel classifiers such as Bayes classifier, linear regression classifier and quadratic discriminant classifier to segment candidate objects. The shape features such as Fourier descriptors, compactness, eccentricity and color features were used and the dimensionality of them was reduced by Fisher transformation. Then they evaluated the performance of a number of classifiers such as k-nearest neighbor classifier, probabilistic neural networks and support vector machine. The experimental results indicated that the sensitivity and specificity of all tested classifiers were above 95%.

Osman et al. [12] have conducted much research on the tissue slides rather than the sputum smears slides for this work. They [13] employed a global thresholding in the hue channel to extract the objects from background. Then an adaptive pixel segmentation based on k-means clustering was applied in the saturation channel derived by CY color model. The undesired candidates were removed by a method of automatic seeded region growing. The final results were refined by a local adaptive thresholding and median filter. They [14] trained a single hidden layer feedforward neural network (SLFN) classifier with six adaptive pixel segmentation based on k-means clustering using the Gaussian weighted adaptive threshold segmentation threshold transform, which can be used in the images with different background colors. In Sect. 2 the candidate objects are extracted based on a marker-based watershed transform, which can be used in the images with different background colors. In Sect. 3 we present the algorithm and implementation based on ASM for the MTB objects recognition. The experimental results and discussion of the proposed algorithm are shown in Sect. 4. The conclusion is given in Sect. 5.

2. Segmentation Using Marker-Based Watershed Transform

Obtaining automatic markers is the key step in marker-based watershed transform. The automatic markers are obtained by using the Gaussian weighted adaptive threshold segmentation and searching the local minimum points in the gradient image. The Gaussian weighted adaptive threshold segmentation is given by

\[ T(x, y) = p(x, y) \otimes G(x, y, \sigma) - b \]  (1)

\[ G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \]  (2)

\[ B(x, y) = \begin{cases} 
1, & \text{if } p(x, y) < T(x, y); \\
0, & \text{else}.
\end{cases} \]  (3)

where \( p(x, y) \) is the pixel value of pixel \((x, y)\) in the gray space, and \( T(x, y) \) is the corresponding segmentation threshold; \( G(x, y, \sigma) \) represents the Gaussian kernel function, and \( \sigma \) determines the size of Gaussian kernel window; \( b \) is an appropriate constant. \( B(x, y) \) is a binary image and its foreground pixels are considered as the initial markers. According to the size of MTB objects, \( \sigma \) is empirically set at 27. The size of the object initial markers is usually small because some objects are low contrast and weak edge. In this case, the objects may be lost after the watershed transform. The local minimum points in the gradient image are considered as the catchment basins to form closed regions during the watershed transform. Therefore, the local minimum points close to the color traits of the MTB objects can be added to the initial markers to increase the numbers of the segmented regions. The local minimum points are searched in a neighborhood of 12 pixels around the initial markers in the gradient image. To avoid the over segmentation, the searching process is only applied to the area within 60 pixels from the initial markers.

After the stage of initial segmentation using the marker-based watershed transform, the segmented regions are merged using the criterion of maximum similarity of adjacent regions to avoid the over segmentation. The color histogram of the region is computed in the RGB color space for similarity measure as

\[ \rho(R, Q) = \sum_{u=1}^{N} \sqrt{\text{hist}_R^u \cdot \text{hist}_Q^u} \]  (4)

where \( \rho(R, Q) \) is the Bhattacharyya coefficient ranging from 0 to 1; \( \text{hist}_R^u \) and \( \text{hist}_Q^u \) denote the normalized histogram of the two regions \( R \) and \( Q \) respectively, and \( u \) denotes the element of the histogram. \( N \) denotes the dimension of the histogram. Finally, multi-thresholds segmentation including background classification and color traits analysis of the MTB objects are utilized to eliminate the impurities [15].

3. Recognition Based on ASM

The shape feature of the MTB is an important measure for
recognition. The MTB is a non-rigid object with a beaded appearance, and the shapes in different MTB objects vary. The MTB objects present differently in length, width and curvature, as shown in Fig. 1.

ASM was first proposed by Cootes et al. [16], [17] and has played an important role in face recognition. The principle of ASM is establishing a statistical shape model which can represent the shape variation of the objects by a set of the landmarks extracted from different object shapes. Thus, ASM is an excellent method for describing the deformation regulation of the object shapes, especially for the deformable objects. Therefore, the ASM is used in our method to model the shape variation of the MTB objects. Assuming that the object shape is represented by average component and variant component as

\[ s = \bar{s} + \Delta s \]  

(5)

where \( s \) denotes the object shape vector, composed of average component \( \bar{s} \) and variant component \( \Delta s \). Each MTB object can be represented by this model.

3.1 Point Distribution Model of MTB

Rational landmarks are important for modeling the shape variation of the objects. The approximate contour and shape characteristics of an object should be represented by these landmarks. In the training process of face recognition based on ASM, the selection of the landmarks is the key step. Since the MTB presents a beaded appearance, the landmarks can be automatically selected from the object boundary to describe the object shape. The single-pixel wide object skeleton is used to find the landmarks. The skeleton is extracted using the thinning method based on the mathematical morphology. The thinning process can be defined by the hit or miss transform

\[ A \otimes B^i = A - (A \otimes B^i) \]  

(6)

where \( \otimes \) denotes the thinning operation, and \( \otimes \) denotes the hit or miss transform. \( B^i \) is the structure element sequence with \( i \in [1, 8] \). The thinning process of the set \( A \) is successively performed by the whole structure element sequence until the generated result will not change.

3.1.1 Key Points of Skeleton

To find more landmarks, the skeleton is equally divided into eight parts by nine key points (including two endpoints, in Fig. 2 (a)). The slopes of these key points except the endpoints are calculated through the quadratic curve fitting based on the least squares error method. The fitting points are composed of the ones between two adjacent key points of the key point. If the segmented boundary at the locations of the ends is not smooth, the skeleton will be also affected to some extent. So for the endpoints, the slope is directly calculated by its adjacent key point to avoid the errors caused by segmentation. The tangential and normal directions of the key points on the skeleton can be obtained from their slopes. The landmarks can be found through the tangential and normal directions of the key points.

3.1.2 Landmarks of Contour

Each key point has two opposite normal directions, and the landmarks can be discovered by extending the key points to the object boundary along the normal directions. The intersection points between the extension lines and object boundary are the landmarks. In order to improve the accuracy and the adaptability for different object shapes, the extending directions are increased from usual four- or eight-neighbors to 16-neighbors, which is shown in Fig. 2 (b). The computed tangential and normal directions are quantized to these 16 directions.

The two ends of MTB objects often have high curvature and big variation in shape. Therefore, more landmarks are needed to characterize the two ends. Since the endpoints of the skeleton have distances to the ends of object boundary, which is shown in Fig. 2 (a), six more landmarks are added as follows. By extending the two endpoints to the object boundary along their tangential directions respectively, we add two intersection points (points 1 and 13 in
the same coordinate system before statistical analysis. The shape vectors of the training set are aligned to remove the effects of translation, rotation and scale transformation from different coordinate systems.

To align two shape vectors, \( s_1 \) and \( s_2 \), the transformation \( T \) is defined as

\[
T(s_2) = R(s, \theta) \cdot s_2 + t
\]

where \( R(s, \theta) \) denotes the rotation matrix with scale factor \( s \) and rotation angle \( \theta \); \( t \) denotes the translation vector. Then the optimal parameters of this transformation can be calculated by the least squares method as follows

\[
(s, \theta, t) = \arg \min[(s_1 - T(s_2))^T(s_1 - T(s_2))]
\]

To align a number of shape vectors, the popular algorithm [18] is Generalized Procrustes Analysis (GPA), which aligns all the shape vectors to the mean vector via iterative update. This process will be terminated until the sum of the distances of each shape to the mean shape is minimized. It is evident that each landmark of the shape has different weights in describing the shape variation. The landmark is more stable if the weight is higher. A weight matrix \( W \) is introduced in Eq. (9) to decrease the aligning errors as follows

\[
(s, \theta, t) = \arg \min[(s_1 - T(s_2))^T W(s_1 - T(s_2))]
\]

where \( W \) is a diagonal matrix composed of landmarks weights, and \( w_i = w_{i+n} \) because they belong to the same landmark.

\[
W = \text{diag}(w_1, w_2, \ldots, w_{2n-1}, w_{2n})
\]

Considering a training set \( \{ s_1, s_2, \ldots, s_{m-1}, s_m \} \), we compute the distances between landmarks within \( s_i \), \( i \in [1, m] \). The result is denoted as \( D_i(j, g) \), \( j, g = 1, 2, \ldots, n - 1 \). \( j \) and \( g \) denote the index of the landmarks. The variance of \( D_i(j, g) \), denoted as \( V(j, g) \), is used to indicate the variation degree of the landmarks. Then the weight element \( w_i (i \in [1, n]) \) is given by

\[
w_i = \frac{1}{\sum_{g=1}^{n} V(j, g)}
\]

Figure 4 represents the normalized weights of the landmarks. From the values of the weights in different landmarks we observe that the primary deformation on the shapes occur at the locations of the object ends for their low weights.

3.3 Modeling Shape Variation

After the shape aligning, the principal component analysis (PCA) is utilized to reduce the data dimensionality of the training set and to find the regulation of shape variation of the MTB objects. The main step of PCA is as follows.

1. Compute the mean vector of the training set,
data is retained, which fulfills the requirements of PCA.

The landmarks approaching the ends of the objects have low weights. It means that the shapes between objects have obvious and frequent deformation in these landmarks.

Table 1

| Number | Eigenvalue | Proportion | Total Proportion |
|--------|------------|------------|------------------|
| 1      | 380.38     | 58.52%     | 58.52%           |
| 2      | 89.27      | 13.73%     | 72.25%           |
| 3      | 44.87      | 6.90%      | 79.15%           |
| 4      | 31.35      | 4.82%      | 83.97%           |
| 5      | 27.14      | 4.17%      | 88.14%           |
| 6      | 17.87      | 2.75%      | 90.89%           |
| 7      | 11.92      | 1.83%      | 92.72%           |
| 8      | 6.00       | 0.92%      | 93.64%           |
| 9      | 4.66       | 0.71%      | 94.35%           |
| 10     | 4.20       | 0.64%      | 94.99%           |

\[
\bar{s} = \frac{1}{m} \sum_{i=1}^{m} s_i
\]  

2. Compute the covariance matrix of the training set,

\[
\Sigma = \frac{1}{m-1} \sum_{i=1}^{m} (s_i - \bar{s})(s_i - \bar{s})^T
\]  

3. Compute the eigenvectors and eigenvalues of the covariance matrix.

\[
\Sigma \cdot v_i = \lambda_i \cdot v_i
\]

The higher is the eigenvalue, the more dramatic variation produced along the directions of the corresponding eigenvector. We choose largest \(k\) eigenvalues \(\lambda_k (\lambda_k \geq \lambda_{k+1})\) and their corresponding eigenvectors \(v_k\) as the principal components. The values and contribution ratio of the largest \(k = 10\) eigenvalues calculated from the used training set are presented in Table 1. The number of the primary components depends on their contribution ratio \(p_{con}\) and accumulative contribution ratio \(\eta\) in the all eigenvalues which are corresponding to the proportion and total proportion in Table 1 respectively. We choose the eigenvalues of which \(p_{con} > 1\%\) as the primary components. Then the largest \(k = 7\) primary components are chosen and the accumulative contribution ratio \(\eta = 92.72\%\) indicates that most information of the raw data is retained, which fulfills the requirements of PCA.

The matrix \(X\) is composed of eigenvector \(v_k\), where \(X = [v_1, v_2, \ldots, v_k, v_l]\). The eigenvectors are orthogonal, so any element in the space which is composed of these eigenvectors can be uniquely expressed. Thus the variant vector \(\Delta s\) can be approximated as

\[
\Delta s \approx X \cdot b
\]

where \(b = [b_1, b_2, \ldots, b_{k-1}, b_k]\) is the \(k\) dimensional shape parameter vector. \(b\) can be adopted to control the shape variation. From Eq. (5) the shape vector can be approximately given by

\[
s \approx \bar{s} + X \cdot b
\]

where \(\bar{s}\) is the mean vector computed in the training process. Using Eq. (17), any appropriate object shape can be generated by varying the elements of \(b\) as shown in Fig. 5.

In Fig. 5, the first principal component represents the ordinary curving variation of the MTB; the second and the third ones represent the variation in length and width respectively; the forth principal component represents the S-type curving variation; the other 3 components represent some local curving variation of the objects. The MTB are non-rigid rod objects, and the primary shape variation of the MTB objects are in terms of ordinary curvature, length, and width and S-type curvature, which can be described by the components of parameter \(b\).

3.4 Recognition Rules Based on ASM

From Eq. (17) the parameter vector is given by

\[
b = X^T (s - \bar{s})
\]

where \(X^T\) and \(\bar{s}\) are known after the training stage, and the shape vector \(s\) can be obtained by Eq. (7). Therefore, the

![Fig. 4](image.png)

**Fig. 4** The normalized weights of the landmarks. The serial numbers are consistent with Fig. 2 (a). The landmarks approaching the ends of the objects have low weights. It means that the shapes between objects have obvious and frequent deformation in these landmarks.

![Fig. 5](image.png)

**Fig. 5** Deformation regulation of the shape model. Each row shows the deformation regulation for each principal component while other components equal to 0. And each column shows its different variation degree and orientation. From left to right the variation are \(-5 \sqrt{\lambda_i}, -3 \sqrt{\lambda_i}, -1.5 \sqrt{\lambda_i}, 0, 1.5 \sqrt{\lambda_i}, 3 \sqrt{\lambda_i}, 5 \sqrt{\lambda_i}\), and \(\lambda_i\) is the eigenvalue and \(i\) is the serial number of the principal component. The middle column is the mean shape of the MTB objects.
parameter vector $b$ of the candidate object can be calculated from Eq. (18). The value of $b$ is limited by the shape model. The characteristic parameters can be used to recognize whether the object is the MTB or not.

### 3.4.1 Parameter Constraint

From the definition in Eq. (18), the higher are the absolute values of the elements $b_i$, the larger is the variation of the shapes. We assume that the elements of $b$ follow the Gaussian distribution given by

$$p(b) = \frac{1}{(2\pi)^{k/2}\Sigma_k^{1/2}} \exp\left\{-\frac{1}{2} b^T \Sigma_k^{-1} b\right\}$$  \hspace{1cm} (19)$$

where $k = 7$ is the number of the principal components, and the mean vector equals to $0$. The covariance matrix $\Sigma_k = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_5)$, where $\lambda_i$ ($i \in [1, k]$) are the first $k$ eigenvalues. Then the range of the element $b_i$ is $\pm 3 \sqrt{\lambda_i}$ to sustain the object shape accepted, and $\lambda_i$ is the standard deviation. To improve the robustness in recognizing and decreasing the effects of segmentation errors, the range of $b_1$ except $b_2$ and $b_3$ can be enlarged to $\pm 5 \sqrt{\lambda_i}$ according to the deformation regulation of each principal component shown in Fig. 5. $b_2$ and $b_3$ represent the variation of the MTB objects in length and width respectively, which are not the same as the other components. The variation of the other components is symmetric in positive and negative directions. For example, the first row in Fig. 5 shows the shape variation of the first component $b_1$ with varying degrees. The shapes corresponding to $b_1 = -5 \sqrt{\lambda_i}$ and $b_1 = 5 \sqrt{\lambda_i}$ are symmetric while $b_2$ and $b_3$ are not. The variation of size is relevant to $b_2$ and $b_3$. If their absolute values are too large, the size of the object shape will not be accepted. From Fig. 5 we can see that if $b_2 = -5 \sqrt{\lambda_i}$, the generated shape almost becomes a line. Therefore, the range of $b_2$ and $b_3$ is $\pm 4 \sqrt{\lambda_i}$ to decrease the false rate. To sum up, the range of the parameter vector $b$ of the MTB objects should satisfy

$$\left\{ \begin{array}{l} |b_i| \leq 5 \sqrt{\lambda_i}, \ (i \neq 2, 3) \\ |b_i| \leq 4 \sqrt{\lambda_i}, \ (i = 2, 3) \end{array} \right.$$  \hspace{1cm} (20)$$

$|b_i|$ is the element of shape parameter vector $b$ and $i$ is an integer over the range 1 to 7; $\lambda_i$ is the eigenvalue of each principal component in Eq. (15).

### 3.4.2 Width Consistency Constraint

During the process of searching landmarks, two landmarks are found by expending each key point on the skeleton to the object boundary along their normal directions. The Euclidean distances of these landmarks to their corresponding key points change a little for the MTB objects, denoted as $d_{mi}, m = 1, 2, 3, \ldots, 18$. This simple and effective feature can be introduced for the MTB objects recognition. We hence call this feature width consistency constraint, denoted by the variance of $d_{mi}$. As a result the normalized variance of the distances, denoted as $\nu_{d_{mi}}$, is added to measure the changes of object width along the skeleton. The variances of the MTB objects are small; an appropriate threshold $\nu_{thr}$ is set according to the statistical result. The width consistency constraint of the MTB objects is given by

$$\nu_{d_{mi}} \leq \nu_{thr}$$  \hspace{1cm} (21)$$

According to Eqs. (20) and (21), given the shape parameter vector $b$ and the width consistency constraint $\nu_{d_{mi}}$ of the candidate objects, the candidate objects can be recognized as the MTB objects if they satisfy the following conditions

$$\left\{ \begin{array}{l} |b_i| \leq 5 \sqrt{\lambda_i}, \ (i \neq 2, 3) \\ |b_i| \leq 4 \sqrt{\lambda_i}, \ (i = 2, 3) \end{array} \right.$$  \hspace{1cm} (22)$$

### 4. Experimental Results and Discussion

The sputum smear slides were collected at the Provincial Tuberculosis Hospital in Changsha, Hunan, China. The images of these slides are captured by a CCD camera mounted on the light microscope with a 100x oil objective. The test images are 24-bit color images with a resolution of 680x512 pixels.

#### 4.1 Segmentation Results

The segmentation results of different cases are shown in Fig. 6. Good segmentation results are obtained in the images with different background colors such as white, blue, and red. The low-contrast MTB objects in Figs. 6(c) and (e) are successfully extracted in Figs. 6(d) and (f). Though the edge of the MTB objects in Fig. 6(g) is fuzzy, caused by the effect of focus in image capturing, the objects are still integrally segmented in Fig. 6(h). One of the MTB objects in Fig. 6(i) is quite inconspicuous. This object was lost in Fig. 6(i) for its great variation to the color characteristics of the normal objects. Some impurities whose color characteristics approach the MTB objects’ are retained after the segmentation, as shown in Figs. 6(k) and (l).

#### 4.1.1 Quantitative Evaluation of Segmentation

The segmentation algorithm was tested by 50 images. The accuracy is assessed by Zijdenbos similarity index (ZSI) coefficient [19]

$$ZSI = 2 \frac{|A_1 \cap A_2|}{|A_1| + |A_2|}$$  \hspace{1cm} (23)$$

where $A_1$ and $A_2$ denote the areas of objects generated by the segmentation algorithm and ground truth respectively, $\mu_{ZSI}$ and $\sigma_{ZSI}$ denote the mean value and standard deviation of $ZSI$ respectively. The evaluation result of the proposed algorithm is $\mu_{ZSI} = 0.8384$ and $\sigma_{ZSI} = 0.04$, which indicate that the algorithm has well segmentation accuracy.
Fig. 6  Segmentation results of the MTB images. (a), (c) and (e) are the positive images with different background color: white, blue and red. (g) is the image out of focus. (i) and (k) are other examples of white background images, and (k) is the negative image with no MTB. (b), (d), (f), (h), (j), and (l) are the segmentation results of (a), (c), (e), (g), (i), and (k) respectively.

4.2 Recognition Results

The recognition results of the images in Fig. 6 are shown in Fig. 7. The bacillus objects with different shapes are detected successfully, and some objects with low contrast and weak edge (such as in Figs. 7(b) and (c)) are recognized. The MTB objects in the out-of-focus image are also recognized in Fig. 7(d). However, one of the MTB object is lost in Fig. 7(e) due to the effect of segmentation result. Some false objects are labeled as the MTB objects in Fig. 7(f) because both their colors and shapes are very close to the MTB objects'. It is difficult to recognize them accurately from some special images even for laboratory technicians.

4.2.1 Quantitative Evaluation of Recognition

The performance of our algorithm is evaluated quantitatively in terms of two levels: object level and image level. Statistical measures, sensitivity, specificity and accuracy, are applied in this work given by

\[
\text{sensitivity} = \frac{TP}{TP + FN} \times 100\% \\
\text{specificity} = \frac{TN}{TN + FP} \times 100\% \\
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{24}
\]
where $TP$, $FP$, $TN$ and $FN$ denote true positive, false positive, true negative and false negative objects respectively. The statistical variables such as $TP$ and $FP$ are obtained by comparing the detected results with the ground truth, which was given by expert pathologists. For the image level evaluation, if there is no positive object detected in one image, the image will be regarded as the negative one.

The experimental results of our approach are compared with the decision tree [6] and intersection kernel support vector machines (IKSVMs) [8]. 1000 objects are provided for the training set including 500 positive objects and 500 negative objects. The two classifiers decision tree and IKSVMs are trained by the training set and evaluated by 240 test images including 160 positive images and 80 negative images. For the decision tree classifier, we characterize each object using compactness and the first four Hu moments [6]. The structure of the trained decision tree is optimized via the cross validation method with an accurate rate of 95.71%. For the IKSVMs, we characterize each object using eight Hu moments, fourteen geometric and photometric features, and a hundred and thirty-six histograms of oriented gradients (HOG). The HOG features are extracted from each 64x64 patch using two scales (64x64 and 16x16) and eight orientations. Due to the difference of image resolution, the size of each patch is 24x24 to extract eighty HOG features [8]. Thus, we obtain a higher dimensional feature vector, 158 dimensions, for representing each candidate object in the IKSVMs.

The quantitative evaluation results of the three methods are shown in Table 2, and ‘Sen’, ‘Spe’, ‘Acc’ denote the sensitivity, specificity and accuracy respectively. We can see that our method is better than the other two methods, especially in specificity. The decision tree classifier using just 5-dimensional features performs better than the IKSVMs using 158-dimensional features. The IKSVMs with multi-dimensional features performs well in the fluorescence microscopy images [8] but not well in the conventional microscopy images. The specificity of the IKSVMs is too low in both two levels, which means that it has a high false alarm rate. The backgrounds of the conventional microscopy images are more complicated because there are various impurities in these images. The 158-dimensional features used in the IKSVMs have some redundancies, and they are not sensitive enough to the impurities. Some impurities that have similar colors and shapes to the MTB objects would be classified as positive objects.

| Object | Method   | Sen (%) | Spe (%) | Acc (%) |
|--------|----------|---------|---------|---------|
| Our method | 90.71    | 97.34   | 92.37   |
| Decision tree | 89.45    | 93.22   | 90.39   |
| IKSVMs | 89.69    | 66.87   | 83.97   |

| Image   | Method   | Sen (%) | Spe (%) | Acc (%) |
|---------|----------|---------|---------|---------|
| Our method | 99.57    | 95.00   | 97.91   |
| Decision tree | 98.75    | 86.25   | 94.58   |
| IKSVMs | 98.13    | 42.5    | 79.58   |

Table 2: Quantitative evaluation in two levels.

5. Conclusion

This paper employed the marker-based watershed transform to extract the candidate objects in the images captured from ZN-stained sputum smear slides. For the MTB object recognition, the shape model based on the ASM was first introduced in this work. The MTB objects are more accurately characterized using the trained shape model than the commonly used geometric features and shape descriptors. In the process of shape modeling, a point distribution model of the MTB objects was proposed to extract the landmarks. The parameter constraint and width consistency constraint were adopted to classify the single and non-bacillus objects. Experimental results have demonstrated the effectiveness of the proposed algorithm. The touching-bacillus objects are composed of several connected single-bacillus objects. These objects cannot be recognized by the proposed algorithm, since the shape model of the MTB objects was trained from the single-bacillus objects. Therefore, this might have some impact on the statistical results. The big challenge to overcome this problem is the segmentation of the touching-bacillus objects. To the best of our knowledge, there was no related work for the segmentation of the touching-bacillus objects. If the touching-bacillus objects can be segmented separately, the single-bacillus objects embedded in the touching-bacillus objects will be recognized by our approach. The established shape model in the proposed approach can be used in segmenting the touching-bacillus objects. We will look into the research about the segmentation of the touching-bacillus objects in future work.

Acknowledgments

The authors would like to thank doctor Wenbo Zhou at the department of laboratory in the Provincial Tuberculosis Hospital in Changsha, Hunan, China, for providing the sputum smear slides and the ground truth guidance. This work was supported by the National Natural Science Foundation of China under Grant No. 61375032.

References

[1] Global Tuberculosis Report 2014, World Health Organization, 2014.
[2] K. Veropoulos, C. Campbell, and G. Learmonth, “Image processing and neural computing used in the diagnosis of tuberculosis,” Proc. IEE Colloquium on Intelligent Methods in Healthcare and Medical Applications (Digest no.1998/514), pp.8/1–8/4, York, UK, 1998.
[3] M.G. Forero-Vargas, E.L. Sierra-Ballen, J. Alvarez-Borrego, J.L. Pech-Pacheco, G. Cristobal-Perez, L. Alcala, M. Desco, B. Javidi, and D. Psaltis, “Automatic sputum color image segmentation for tuberculosis diagnosis,” Proc. SPIE, vol.4471, Algorithms and Systems for Optical Information Processing V, San Diego, pp.251–261, 2001.
[4] M.G. Forero-Vargas, “Fuzzy thresholding and histogram analysis,” Fuzzy filters for image processing, Series Studies in Fuzziness and Soft Computing, vol.122, pp.129–152, Springer-Verlag, Berlin, Heidelberg, 2002.
A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer, “Statistical models of appearance,” Med. Image Comput. and Computer-Assisted Intervention, MICCAI 2012, Lecture Notes in Computer Science, vol.7512, pp.345–352, Springer-Verlag, Berlin, Heidelberg, 2012.

M.G. Forero-Vargas, F. Sroubek, J. Alvarez-Borrego, N. Malpica, G. Cristobal, A. Santos, L. Alcala, M. Desco, L. Cohen, K.M. Iltekhaaruddin, and A.A.S. Awwal, “Segmentation, autofocusung and signature extraction of tuberculosis sputum images,” Proc. SPIE, vol.4788, Photonic Devices and Algorithms for Computing IV, pp.171–182, Seattle, WA, 2002.

V. Makkapati, R. Agrawal, and R. Acharya, “Segmentation and classification of tuberculosis bacilli from ZN-stained sputum smear images,” Proc. 5th Annual IEEE Conference on Automation Science and Engineering, pp.217–220, Bangalore, India, 2009.

Chao Xu received the B.Sc. degree in electronic information science and technology from Shandong University, Jinan, China, in 2009, and the M.Eng. degree in information and communication engineering from National University of Defense Technology, Changsha, China, in 2011. Now he is pursuing his Ph.D. degree at Intelligent Information Processing Lab in National University of Defense Technology. His research interests include image processing and pattern recognition.

Tao Guan received the B.Eng., M.Eng. and Ph.D. degrees in information and communication engineering from National University of Defense Technology, Changsha, China, in 2007, 2009 and 2015, respectively. His research interests include medical image processing and pattern recognition.

Yongping Zhai received the B.Eng., M.Eng. and Ph.D. degrees in information and communication engineering from National University of Defense Technology, Changsha, China, in 2004, 2007 and 2013, respectively. Now he is a Post-doctoral researcher at Multimedia Information System Lab in National University of Defense Technology. His research interests include computational photography, image processing and pattern recognition.
Yunhui Liu received the B.Eng. degree in applied dynamics from Beijing Institute of Technology, Beijing, China in 1985, the M.Eng. degree in mechanical engineering from Osaka University, Osaka, Japan in 1989, and the Ph.D. degree in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan in 1992. He worked with the Electrotechnical Laboratory, MITI, Japan from 1992 to 1995. Since February 1995 he has been with The Chinese University of Hong Kong and is currently a Professor with the Department of Mechanical and Automation Engineering. He is also an Adjunct Professor of the National University of Defense Technology. His research interests include visual servoing, medical robotics, mobile robots, sensor networks, and machine intelligence. Dr. Liu has published over 100 papers in refereed journals and refereed conference proceedings. He has received numerous research awards from international journals and international conferences in robotics and automation and government agencies. He was an Associate Editor of the IEEE Transactions on Robotics and Automation from 2001 to 2004. He was the general chair of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).