TB-v1: Software toolbox environment for the detection of TB bacilli in auramine staining

Manuel G. Forero Vargas y Gabriel Cristóbal
Instituto de Optica (CSIC), Serrano 121, 28006 Madrid (Spain)
Email: gabriel@optica.csic.es
June 8th, 2004

Abstract

Tuberculosis and other mycobacteriosis are serious illnesses which control is based primarily in early diagnosis. In addition to clinical suspicion, a fundamental support in diagnosis is the staining of clinical samples with gender-specific colors. However, these techniques are not very sensitive so many times we must await the results of culture, which can take up to 2 months. We have developed a software called TB-v1 to help you through the application of techniques of digital image processing in specific staining of mycobacteria, providing specificity rates of around 98% for a sensitivity of 94% which justifies the conclusion that such system is susceptible to clinical applicability. The analysis techniques are based on the extraction of invariant descriptors from the shape of objects in conjunction with color information. The decisions taken by the program is based on Bayesian inference techniques. TB-v1 is a program that analyzes the images that have been captured by the camera attached to a fluorescence microscope and automatically indicates the presence or absence of bacilli in the image.

1. Background and state of the art

Tuberculosis is a communicable disease, predictable and curable, so a key point in its control is rapid diagnosis. The slow growth of mycobacteria makes the diagnosis based on culture may be delayed up to two months. That is why methods based on staining are essential in this type of pathogen. The special structure rich in mycolic acids of mycobacteria makes specific stains exist for these organisms.

Among the specific staining for mycobacteria, auramine stain is currently the most widely used due to its higher sensitivity and ease of observation. For observation using UV microscopy an objective of X20 is used. The estimated time to rule out as negative auramine stain is about 2 minutes. Tuberculosis is a disease that annually affects about 10 million people of which about 3 million die from the disease directly. In Madrid (Spain) the incidence of the disease has gone from as high as 40 cases per 100,000 inhabitants in the mid-nineties to the current estimate of 23 cases of tuberculosis per 100,000 inhabitants and a mortality of 1,5 cases (1999). Despite the large decrease in the incidence of tuberculosis in Madrid, it remains one of the highest rates of tuberculosis in Europe.

There is no doubt that early diagnosis is the first weapon to control a communicable disease like tuberculosis. The increase in the number of people with inmnunosupresion has led to a rise in recent years several cases of mycobacterial tuberculosis which early diagnosis is also essential. Since the major clinically important mycobacteria (M.

tuberculosis complex, M. avium complex, M. kansassii) slow to grow in culture media from 1 to 8 weeks, the specific staining of mycobacteria have always been an essential diagnostic method in any mycobacteriology laboratory.

Since more than a century ago Robert Koch described the first method of staining of mycobacteria, many techniques have been used for that purpose. From the point of view of the microscopy, two types of stains, non-fluorescent and those that are. Among the non-fluorescent staining, the stain most often used is, without doubt, the Ziehl-Neelsen. This staining technique used for coloring the basic fuchsin, the sulfuric acid to 25% for bleaching, and the methylene blue counterstain. Mycobacteria are shown colored red while the background is a bluish color. For viewing you must use the immersion objective (1000 ×) and 300 microscopic fields to be observed, equivalent to three lengths of a theoretical sample button 10x20 mm blade. In general, the time needed to rule out as negative preparation using this technique is about 10 minutes.

State of the art.

Veropoulos [20] developed a method of identification of mycobacteria by observing them first for low magnification (x100) and higher increases (X400), making the acquisition for even larger increases (X630). However, the method described is based on the analysis for medium magnification (x200) given that such magnification is used in clinical practice.
In a first phase, 397 negative images were taken from 31 patients and 75 positive from 4 patients. Afterwards, this set of samples was expanded in order to apply a decision-making procedure based on Bayesian inference. For each patient, between 8-100 RGB images (1600x1200) were collected. For the increases shown (x250), bacilli occupy a small fraction of the total image area (around a factor 6.7 x10-5). To ensure ownership bacilli, 110 of them extracted from 15 positive images. The procedure of identification of bacilli (shown schematically in Fig.1) is based on two stages,

- Identification by color. At this stage we performs a thresholding of the green channel in relation to red channel, in order to extract only the regions whose color is characteristic of staining the bacilli
- Identification by shape. From the green channel, we proceed to edge detection using Canny operator. Then apply several morphological operators in order to draw the forms of the bacillus and eliminate candidate regions whose shape is not characteristic of the bacilli.

Both steps are combined together in order to keep only those objects that have been removed by a detector and also check the status of color. Figure 2 shows an example of object extraction in the case of a negative and positive image.
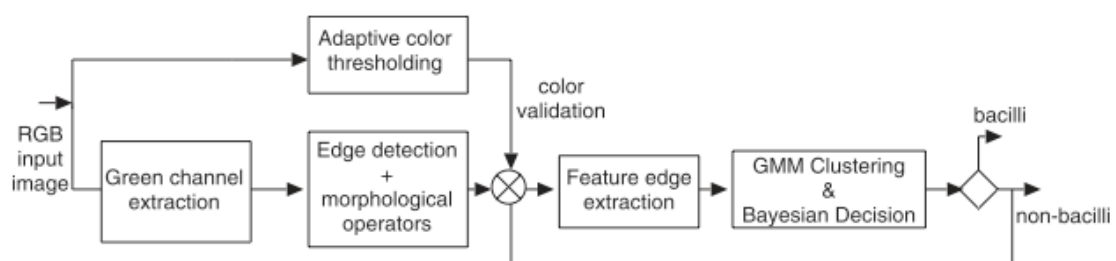
Fig.1. Schematic diagram of system identification of tuberculosis bacilli.

(a)

(b)

(c)

(d)
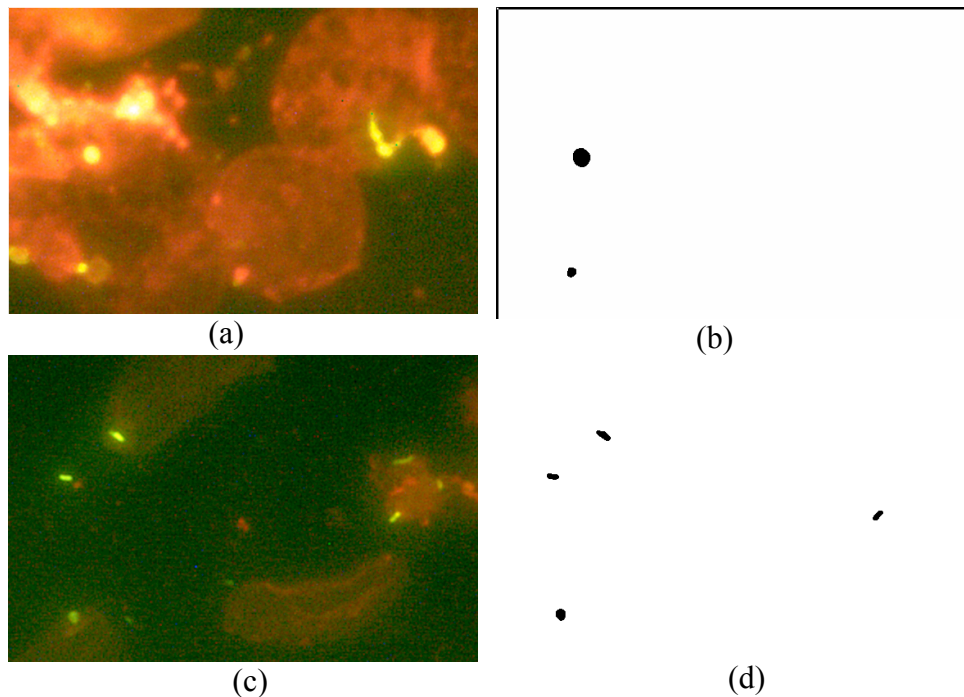
Fig. 2. (a) Example of negative image. (b) Result of segmentation by shape and color. (c) Example of positive (d) Result of segmentation by shape and color. From the images (b) and (d) is made a feature extraction process invariant to geometric transformations. These desciptores can assign each of the detected objects to any class of bacilli or to a kind of rejection.

The bacilli are then detected from a clustering process. To do this, a first training stage of the system from a previously selected group of bacilli is performed. Based on the overall sample of bacilli extracted, we proceed to obtain the centroid of each class. In this particular case, and to study this problem, we found 4 clusters corresponding to 4 different kinds of bacilli. The idenficación of bacilli is then carried through a mechanism based on Bayesian decision theory.

In a second stage of this study, we expanded the database of images of bacilli. Specifically, we considered 38 examinations, 31 of which corresponded to negative reviews and 7 to positive tests. As the number of images, 308 positive imagess were taken which 139 of them were used for the extraction of the bacilli of reference and the remainder i.e. were used 169 to evaluate the method. In relation with negative images a total of 473 were considered. This number is much higher both in terms of the training and testing set allowing to improve the accuracy of the proposed method.
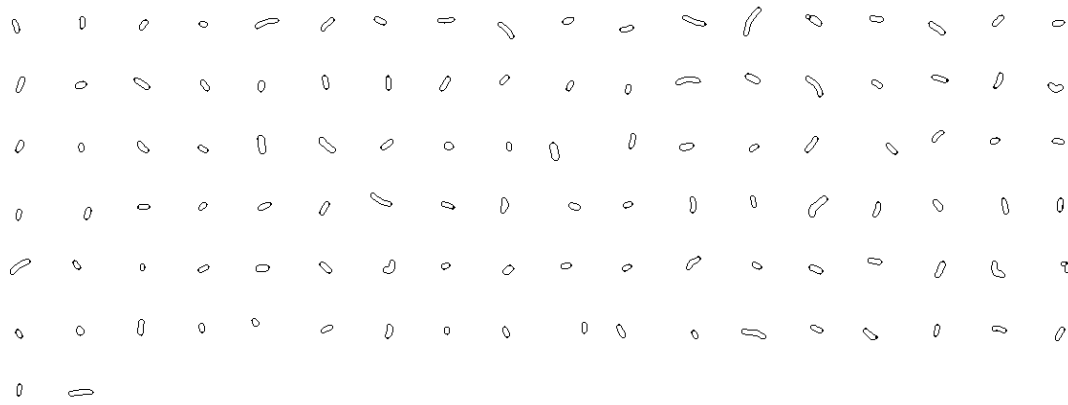
Fig.3 Example of the training bacilli

2. Materials and methods

- A photomicroscope equipped with fluorescent lighting and equipped X20 objective.
- A high-sensitivity digital camera capable of capturing images from the microscope. Examples of suitable cameras for the acquisition of such images are Spot Insight from Diagnostic Instruments and Leica DFC 480.
- A computer to capture images of the microscope, with the IEEE 1394 (Firewire), for connecting the camera. It is also necessary for the execution of the program, Matlab and the Image Processing Toolbox from Mathworks, Inc., ©. It is also necessary NetLab toolbox described below.

3. Results

Table I presents results of diagnostic accuracy level based on the values of specificity and sensitivity. It can be seen that the values obtained support the conclusion that the proposed method provides very good values of diagnostic accuracy, which means it can be concluded that method could be of interest for clinical use.

Table I. The diagnostic accuracy expressed as diagnostic sensitivity and specificity. Specificity is the probability of correctly classifying an individual whose real estate is defined as the negative. Is equal to the result of subtracting one to the fraction of false positives. By contrast, the sensitivity is related to the fraction of true positives. ν represents the decision threshold.

ν	Especificidad	Sensibilidad
0,1	97,89 ± 1,3%	94,67 ± 3,39%
0,05	97,89 ± 1,3%	94,67 ± 3,39%
0,0001	98,10 ± 1,23%	92,9 ± 3,87%

4. References

 [1] Alvarez-Borrego, J., M. C. Chávez-Sánchez y M. A. Bueno-Ibarra. Color correlation
for detecting virus in shrimp. Light for Life and VI Education and Training in Optics and Photonics. Cancún, Quintana Roo, México. JULIO 27-30,1999.

[2] Alvarez-Borrego, J., M. C. Chávez-Sánchez y M. A. Bueno-Ibarra. Detection of virus in shrimp using digital color correlation. 18th Congress ofthe Intemational Commision for Optics, ICO XVIII, Optics for the Next Millennium. AGOSTO 2-6, 1999.

[3] Bishop, P. J., Neumann, G. 1970. The history ofthe ZiehI-Neelsen stain. Tubercle.51:196-206.

[4] Comité de expertos en bacteriología (UICT). 1978. Guía técnica para el diagnóstico de la tuberculosis por microscopía directa. Informes técnicos UICT, suppl. 2.

[5] Dolin, P. J.» M. D. Raviglione, y A. Kochi. 1993. Global tuberculosis incidence andmortality during 1990-2000. Bull. W. H. O. 72:213-220.

[6] M.Forero-Vargas, E. Sierra-Ballen, J. Alvarez-Borrego, J.L. Pech-Pacheco, G. Cristobal, L. Alcala and M. Desco (2001), "Automatic sputum color image segmentation for tuberculosis diagnosis", SPIE 4471, San Diego, CA, 2001

[7] M.Forero-Vargas,F. Sroubek, J. Alvarez-Borrego, N. Malpica, G. Cristobal, A. Santos, L. Alcala, M. Descoand L. Cohen. (2002), Segmentation, autofocusing and signature extraction of tuberculosis sputum images", SPIE 4788, Photonic Devices and Algorithms for Computing, Seattle, WA, 2002

[8] Kubica G. P. 1980. Correlación entre los métodos de tinción y los resultados de los cultivos de micobacterias. Bol. Un. Int. Tuberc. 55:119-126.

[9] Mouriño-Pérez, R. 1999. Optical color correlation for stages recognition of different stages of *Vibrio Cholera* 01. tesis doctoral. CICESE. México.

[10] Mouriño-Pérez, R. R. y J. Alvarez-Borrego. Color correlation for the recognition of different stages Vibrio cholerae 01. Light for Life and VI Education and Training in Optics and Photonics. Cancún, Quintana Roo, México. JULIO 27-30,1999.

[11] Mouriño-Pérez, R. y J. Alvarez-Borrego. Color correlation for the recognition Vibrio

cholerae 01 in seawater. 18th Congress ofthe Intemational Commision for Optics, ICO XVIII, Optics for the Next Millennium. AGOSTO 2-6, 1999.

[12] Pech-Pacheco, J.L. y Alvarez-Borrego, 1998. J. Optical digital system applied to the identification offive phytoplancton species. Mar. Biol., 132: 357-365

[13] Pech-Pacheco, J.L. y Cristóbal, G. 2002, "Automatic slide scanning" en Automatic Diatom Identification (H. du Buf y M. Bayer, eds.), World Scientific, Singapur

[14] Smithwick, R. W., Stratígos, C. 1995. Preparation of acid-fast microscopy smears for profíciency testing and quality control. J. Clin. Microbiol. 110-111.

[15] Unión Internacional contra la tuberculosis. (1978). Guía técnica para el diagnóstico de la tuberculosis por microscopía directa. Bol. Un. Int. Tuberc.

[16] Naumann, D., Helm, D. y Labischinski, H. (1991). "Microbiological characterizations by FT-IR spectroscopy", Nature 351, pp. 81-82

[17] Liu,J., Dazzo, F., Glagoleva, O., Yu, B. y Jain, A.K."CMEIAS: A computer-aided system for the image analysis of bacterial morphotypes in microbial communities"

[18] Kim, T. y Poon, T.(2000), "Three dimensional matching by use of phase-only holographic information and Wigner distribution", *J. Opt. Soc. Am. A*, 17, 12, pp. 2520-2528

[19] B. Schilling, T.-C. Poon, G. Indebetouw, B. Storrie, K. Shinoda, and M. Wu, Three-Dimensional Holographic Fluorescence Microscopy," *Optics Letters*, Vol. 22, 1506-1508 (1997)

[20] K. Veropoulos, G. Learmonth, C. Campbell, B. Knight y J. Simpson, "Automatic identification of tubercle bacilli in sputum. A preliminary investigation", Amalytical and Quantitative Citology and Histology, 21,4, pp. 277-281, 1999

[21] M. Forero-Vargas, E. Sierra, J. Alvarez-Borrego, J.L. Pech-Pacheco, G. Cristóbal, L. Alcala y M. Desco, "Automatic sputum color segmentation for tuberculosis diagnosis", en SPIE Proc. Algorithms and systems for Optical Information Processing, vol. 4471, pp. 251-261, 2001

[22]M. Forero-Vargas, F. Sroubek, J. Alvarez-Borrego, N. Malpica, G. Cristóbal, A. Santos, L. Alcalá, M. Desco y L. Cohen, "Segmentation, autofocusing and signature extraction of tuberculosis sputum images", Proc. SPIE. Photonic Devices and Algorithms for Computing, vol. 4788, pp. 171-182, 2002

[23] M. Forero-Vargas, G. Cristóbal and J. Alvarez-Borrego, "Automatic identification techniques of tuberculosis bacteria", SPIE Proc. Applications of Digital Image Processing XXVI, vol. 5203, pp. 71-81, 2003

[24] M. Forero-Vargas, F. Sroubek y G. Cristóbal, "Identification of tuberculosis bacteria based on shape and color", Real Time Imaging, 10/4, pp. 251-262, 2004.

[25] M. Forero-Vargas, G. Cristóbal and Desco, M., "Automatic identification of Mycobacterium tuberculosis by Gaussian Mixture models", J. Microscopy, 223, pp. 120-132, 2006.

Program TB-v1. User manual.

TB-v1 program in its current version has been implemented as a Matlab toolbox, requiring also the use of NetLab toolbox, described below.

Main Program: TuberculosisMMGDefinitiva.m

DESCRIPTION
This program performs the analysis of images of tuberculosis. For such purpose, this program gets the descriptors of the objects used as training (binary images of individual bacilli). It was assumed that each training image contains only a single bacillus. For the the classification of objects the following descriptors are used Hu1, Hu2, Hu3, Hu11, whereas for the analysis of testing objects the eccentricity and area were used.

1. Program input/output

The flow chart has been replaced by a pseudo-code description of the steps that are carried out by the program.

PROGRAM INPUTS

Training images (individual bacilli binarized) and image to be analyzed. The format of the images is bmp, but image reading may be done in any other format supported by Matlab.

PROGRAM OUTPUTS
ASCII Message screen, indicating the presence of bacilli and the number of bacilli found.

Brief description of the tasks performed by the program:
1. Capture training objects
2. Descriptors of training objects are obtained
3. You get the class centroids
4. Apply Gaussian mixed models for determining the probability density functions of the class of bacilli, consisting of 4 clusters, and determine the rejection thresholds
5. Calculate the covariance matrix of the clusters.
6. Read images to be analyzed
7. Calculate the covariance matrix.
8. Image analyzis:
a. To do this we obtain the candidate objects
b. Descriptors are obtained from the candidate objects. Objects which area and eccentricity are outside the range set of candidate objects are removed
c. A classification by Mahalanobis distance is performed
d. From the distance obtained, it is determined whether or not there is presence of bacilli.

2. List the names of files it contains and auxiliary programs required by the main program.

As mentioned above the main program is called: TuberculosisMMGDefinitiva.m
For its implementation, this program requires the use of the toolbox NetLab:
http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/

Used functions are:
- consist.m - Check that the arguments are consistent
- gmm.m - Generates a Gaussian mixed model with a specific structure
- gmmactiv.m-Computes the activations of a Gaussian mixed model
- gmem.m-EM for Gaussian mixed model
- gmminitmod.m-modified version of the routine gmminit (including Matlab routine kmeans
- gmmpost.m-calculates the probabilities of the classes 'a posteriori ' of a model GMM

Other programs needed

- momentsTBCHuFlusser.m-Hu moment estimation of a binary image
- tuberculosetraining.m-get descriptors of the objects used for training.

The other files are data files needed for training analysis system. The name of these files is a sequence of numbers corresponding to the code of the sample captured in the laboratory. The library is a library uigetfiles.dll Windows that allows you to introduce within the Matlab environment the files mentioned above.

Programming language

The program has been implemented in the Matlab programming language, which is a high-level language widely used by the scientific community.

How to run the program

TuberculosisMMGDefinitiva.m
The main program contains a documentation of each of the steps carried out within the program. This documentation is supplemented with that is contained in this document. To run this program is recommended to decompress all files in a subdirectory, which is included in the path of the programs that will recognize Matlab. Once inside the Matlab environment, the main program TuberculosisMMGDefinitiva is executed, and then the program asks for the training file that are contained in the directory named 929PositivasBinarizadasEditadas, corresponding to 929 bacilli that were edited and processed manually.
Once the system has, the program is now ready to make a decision. The package also includes some examples of positive images (in the directory named

"AuraminasPositivas"), and negative images (in the "AuraminasNegativas") to evaluate the results provided by the program.

COPYRIGHT NOTICE

ACKNOWLEDGEMENTS