A Sputum Smear Microscopy Image Database for Automatic Bacilli Detection in Conventional Microscopy

M. G. F.Costa-*IEEE Member*, C. F. F. Costa Filho-*IEEE Member*, A. Kimura Junior, P. C. Levy, C. M. Xavier and L. B. Fujimoto

***Abstract*— In this work, we present an image database for automatic bacilli detection in sputum smear microscopy. The database comprises two parts. The first one, called the autofocus database, contains 1200 images with resolution of 2816 x 2112 pixels. This database was obtained from 12 slides, with 10 fields per slide. Each stack is composed of 10 images, with the fifth image in focus. The second one, called the segmentation and classification database, contains 120 images with resolution of 2816x2112 pixels. This database was obtained from 12 slices, with 10 fields per slice. In both databases, the images were acquired from fields of slides stained with the standard Kinyoun method. In both databases, accordingly to the background content, the images were classified as belonging to high background content or low background content. In all 120 images of segmentation and classification database, the identified objects were enclosed within a geometric shape by a trained technician. A true bacillus was enclosed in a circle. An agglomerated bacillus was enclosed by a rectangle and a doubtful bacillus (the image focus or geometry does not allow a clear identification of the object) was enclosed by a polygon. These marked objects could be used as a gold standard to calculate the accuracy, sensitivity and specificity of bacilli recognition.**

***Keywords*— Image Database, Microscopy, Detection, Bacilli, focus.**

I. Introduction

For providing an embracing and update assessment of TB epidemic, the World Health Organization has published every year, since 1997, an annual report on global control of tuberculosis.

According to the Global TB control report of 2013 [1], "Tuberculosis (TB) remains a major global health problem. In 2012, an estimated 8.6 million people developed TB and 1.3 million died from the disease (including 320 000 deaths among HIV-positive people). The number of TB deaths is unacceptably large given that most are preventable."

The Millennium Development Goals (MDGs), adopted by healthcare providers, are part of a United Nations Development Programme [2], which provides concrete, numerical benchmarks for tackling extreme poverty in its many dimensions to be achieved by 2015. The program has 8 millennium development goals and 21 targets, measured by 60 indicators. The 6th goal, related to fighting disease epidemics, includes TB. The aim of this goal is "Combat HIV/AIDS, Malaria and other diseases". Within this goal, the following target refers to TB: "Halt and begin to reverse the incidence of malaria and other major diseases". Related to this target, the following indicators refer to TB: halt and begin to reverse TB incidence by 2015; reduce prevalence and deaths of TB by 50% compared to the 1990 baseline.

To achieve these goals the WHO adopted a Partnership Global Plan to Stop TB [2], launched in January 2006. This plan includes smear sputum microscopy as the main diagnostic tool.

For TB diagnostic with sputum smear microscopy, two different technics are usually employed: conventional microscopy and fluorescence microscopy. In the literature, it is reported that Fluorescence microscopy is on average 10% more sensitive than conventional microscopy [3]. Comparing to conventional microscopy, the main drawbacks of fluorescence microscopy are the following: 1) The relatively high costs of the microscopy unit and of its maintenance; 2) An advanced technical skill is required to handling and maintenance of the optical equipment.

The manual screening for bacillus identification has the following drawbacks: a huge variability in sensitivity and a labor-intensive consuming between 40 minutes and 3 hours. Depending on patient's level of infection, it is necessary to analyze 40-100 images [4].

The first automatic methods for bacilli screening were developed for fluorescence microscopy images [5, 6]. In conventional microscopy, the first methods for automatic bacilli screening were published only in 2008 [7]. Other methods for automatic bacilli screening were published in recent years [4, 8, 9, 10].

Steps involved in automated methods include auto focusing, image capture, bacilli segmentation and bacilli

M. G. F. Costa is with the Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas, Manaus, AM, 69077-000, Brazil (e-mail: mcosta@ufam.edu.br).

C. F. F. Costa Filho, is with the Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas from Technology and Information Center, Manaus, AM, 69077-000, Brazil (e-mail: ccosta@ufam.edu.br).

A. Kimura Junior is with the Automation and Control Department, State University of Amazon, Manaus, AM, 69050-020 Brasil, (e-mail: akimurajr@gmail.com).

P. C. Levy s with the Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas from Technology and Information Center, Manaus, AM, 69077-000, Brazil (e-mail: pamela.levy@gmail.com).

C. M. Xavier is with the Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas from Technology and Information Center, Manaus, AM, 69077-000, Brazil (e-mail:clahildek@gmail.com).

L. B. Fujimoto is with the Medical School, Federal University of Amazon, Manaus, 69020-060, Brazil (e-mail: olfujimoto@gmail.com).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

classification. For developing an automatic method, a main concern is the disposal of tuberculosis image databases, both for finding a best auto-focusing measure for TB slices as to finding a method for TB bacilli segmentation and classification.

Regarding auto-focusing, one can imagine that there is a general procedure to find the focus of any type of image. According to Subbaro and Tyan [11], nevertheless, there is no best focus measure that can be used for auto focusing of different image types. The best focus measure could be different for different objects depending on both image content and noise characteristic. Therefore, it is important to find the best focus measure that can be used in TB auto focusing, and with this objective, a TB image database with images in focus and out of focus is essential.

Regarding TB bacilli segmentation and classification, a different image database is needed to train and test segmentation and classifier methods. The main characteristic of this database is that images must all be in focus.

A literature review reveals that all authors have their proprietary bacilli databases. This makes it difficult to search for the best auto-focusing measure and compare different methods for bacilli segmentation and classification.
Table 1 shows a review of certain studies published in the literature, showing the type of work (auto-focusing or bacilli segmentation and classification) and image database.

In this study, we present a database for automatic bacilli detection of TB conventional microscopy images, divided into two sections. The first one suited for autofocusing step and the second one, for bacilli segmentation and classification.

II. METHODS

Image acquisition was accomplished at the Mycobacteriology Laboratory of National Research Institute of Amazonia– INPA. An acquisition station was set up (a digital camera, a light microscope and a computer). The digital camera used was a Canon PowerShot A640, which couples with a 10-megapixel CCD imager sensor with a 4x optical zoom. The spatial resolution was set up to 2816 x 2112 pixels, 24 bits per pixel (RGB images). The microscope used was a Zeiss Axioshop 40. It was employed a magnification of 100x and numerical aperture of 1.25. The PC used for image acquisition had a core 2 duo 2.0 GHz with 3GB RAM. The sputum samples from patients suspected of pulmonary TB cases were prepared by the Kinyoun stain method.

Aiming at developing an automatic method for bacilli detection, two databases were constructed. The first one was for adjusting the best focus metric for autofocus step and the second for segmentation and classification steps. In the sequence, we detail each database.

A. TB Autofocus Database

Three groups of images were generated. In the first and second groups, images were acquired from fields of slides stained with the standard Kinyoun method. In the first group, images were taken from fields featuring high density of background content, while in the second group, images were taken from fields of slides with low-density background content. In the third group, images were acquired from fields of slides stained with the modified Kinyoun method, without being counterstained with methylene blue solution. These groups were created with the aim of identifying the influence the absence of a counterstain and of the density of background content in the focus of the image.

Examples of images from three groups used in the experiment are shown in Fig. 1. In each group, images were obtained from 4 slides. From each slide was selected 10 fields. From each field, 10 samples were acquired with different focal lengths. It was employed a focal length step of 2.5 μm. Each group of ten samples acquired from one field constitute one set of images. The sample in focus was found by a trained operator. We set the sample in focus at image five. The image sharpness decreased moving in the direction toward images one and ten. The total image sets for the experiment was 120 (40 per group) totaling 1200 images.

TABLE I. Details of TB image databases used in automatic bacilli detection

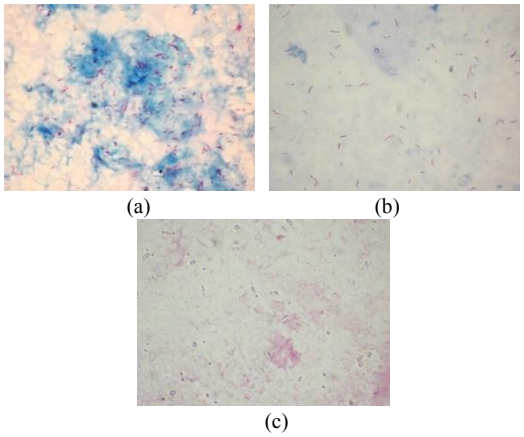
Reference	Type of work	Image Database
[12]	auto-focus	No details about image database. Database not available
[13]	auto-focus	1200 images: resolution of 2816 x 2112 pixels; 12 slides; 10 fields/slide; 10 images/stack. Database published in this work.
[14]	auto-focus	1300 images: resolution of 1392 × 1040 pixels; 13 slides; 5 fields per slide; 20 images/stack. Database not available.
[8]	Bacilli segmentation and classification	resolution: 720x480 pixels; 19 slides; 20 to 100 images per slide.
[9]	Bacilli segmentation and classification	120 images: resolution 2816x2112 pixels; 12 slices; 10 fields/slice. Database published in this work
[15]	Bacilli segmentation	100 images: 10 slides; 10 fields/slice. Database no published

B. TB Segmentation and Classification Database

The database comprises 120 images. Sputum smear microscopy slices of 12 patients (10 fields for each patient) were obtained to build this database.

Depending on the background content, the database is divided in two groups. Group 1 consists of images with high density of background content (HDB), while Group 2 consists of images with low density of background content (LDB). The LDB group is characterized by a weak presence of counterstain with methylene blue solution in the background, while the HDB group is characterized by a strong presence of this same counterstain. Fig. 2 shows one

image of the LDB type and other image of the HDB type. As shown, in the background of HDB images, there is a prevalence of blue color, while in the background of LDB images there is a prevalence of white.



(a) (b)

(c)

Figure 1. Examples of microscopy images from: (a) first group, (b) second group, (c) third group.

For classification of the images in these two groups, the hue component of the HSI space was used. For each image, the percentage of pixels with a hue component in the blue color range, (0.5-0.7), was obtained. Fig. 3 shows this percentage, for images of both groups, represented with bars. The bars were organized so that the last 60 images have bars corresponding to high percentage values and the first 60 images have bars corresponding to low percentage values. An experimental threshold of 13.56 was established to separate images as belonging to group HDB or to group LDB. This threshold value is shown as a horizontal line in Fig. 3. When the bar value was greater than this threshold, the image was considered as belonging to the HDB group. When the bar value was less than this threshold, the image was considered as belonging to the LDB group.

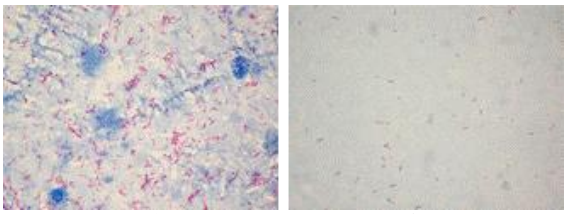
In all the 120 images, the identified objects were enclosed within a geometric shape by a specialist. A true bacillus was enclosed in a circle. An agglomerated bacillus was enclosed by a rectangle and a doubtful bacillus (the image focus or geometry does not allow a clear identification of the object) was enclosed by a polygon. These marked objects could be used as gold standard to calculate the accuracy, sensitivity and specificity of bacilli recognition. In our previous works [9, 10], for calculating these measures, the doubtful bacilli and agglomerated bacilli (because it is not possible to know how many bacilli there are in one agglomeration) were not taken into account. Fig. 4a (LDB image) and Fig. 5b (HDB image) show marked images as described here.

III. Results

The database can be accessed in http://www.tbimages.ufam.edu.br. First is it requested that the user fill out and send a form to authors that authorize the access sending a login and password. In the main page, there is a short description of the database. In the main page, on the tab titled TB_IMAGES_DB_FOCUS_.V1, the user has access to the focus database. In the main page, on the tab titled TB_IMAGES_DB_BACILLI.V1, the user has access to segmentation and classification database.

In a previous study [13], using the autofocus database, we did a systematic analysis of focus functions in conventional sputum smear microscopy for tuberculosis. The main accomplishment of this work was to show that an autofocus function based on variance measures produced the best results for tuberculosis images.

In another study [9], using the segmentation and classification database, we presented a method for bacilli recognition employing neural networks to bacilli segmentation and a color filter, proposed by the authors, for bacilli classification. The result was a sensitivity of 91.53% in bacilli detection.



(a) (b)

Figure. 2. (a) Image with high density of background content. (b) Image with low density of background content.

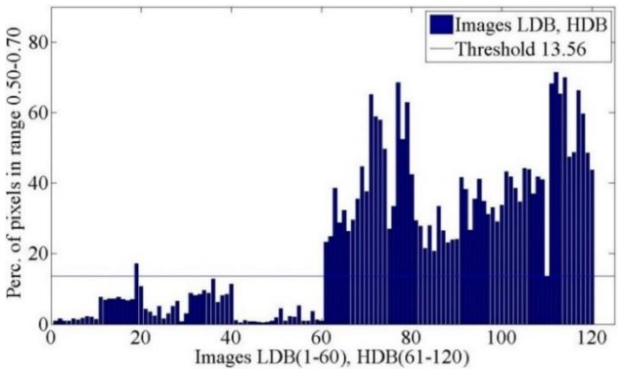


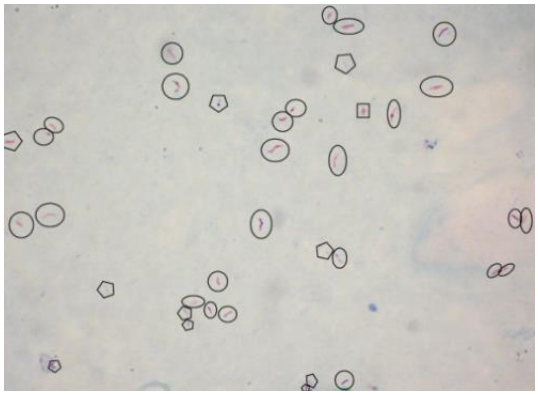
Figure 3. Bar percentages for all images of groups LDB and HDB.

IV. Conclusion

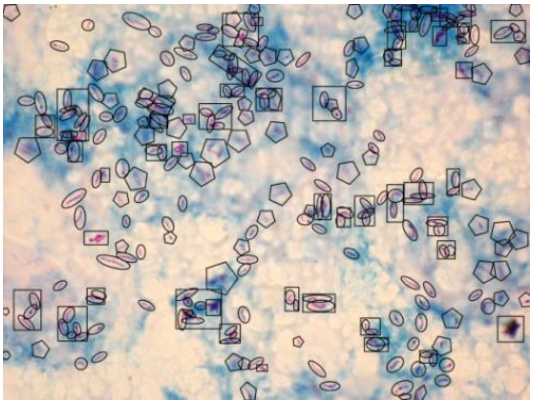
An image database of conventional sputum smear microscopy of tuberculosis patients was presented. Auxiliary images, with objects marked as bacillus, agglomerated bacillus and noise, were generated. Now, we are improving this database, including information that relates a given image with tuberculosis staging.

Acknowledgment

We would like to thank FAPEAM and Samsung for financial support. We also like to thank AcademicEnglishSolutions.com for revising the text.



(a)



(b)

Figure 4. (a) LDB image with marked objects. (b) HDB image with marked objects. Circle – true bacillus; Polygon – doubtful bacillus; rectangle – agglomerated bacilli.

REFERENCES

[1] WHO, Global TB Control report 2013 at http://www.who.int/tb/publications/global_report/2013.

[2] WHO, Global TB Control report 2010 at http://www.who.int/tb/publications/global_report/2010.

[3] K. R . Steingart , M. Henry , V. Ng, P. C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. Perkins, M. A. Aziz and M. Pai, "Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review," Lancet Infection Disease 6, 2006, pp. 570-581.

[4] M. Sotaquirá, L. Rueda and R. Narvaez, "Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis," in *Proc. of International Conf. on Digital Image Processing*, , 2009 pp. 117-121.

[5] K. Veropoulos , C. Campbell, G. Learmonth , B. Knight and J. Simpson, "The automated identification of tuberculose bacilli using image processing and neural computing techniques," in *Proceedings of 8th International Conf. on Artificial Neural Networks 2*, ,1998, pp. 797-802.

[6] M. G. Forero and G. Cristóbal, "Automatic identification techniques of tuberculosis bacteria," *Proc. SPIE 5203* , 2003, pp. 71–81.

[7] M. G. F. Costa, C. F. F. Costa Filho, J.S. Ferreguette, J. Salen and M. O. Lima, "Automatic identification of mycobacterium tuberculosis with conventional light microscopy," *30th Annual Int. Conf. of the IEEE Eng in Med and Biol Society 1* ,2008, pp. 382-385.

[8] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth and T. S. Douglas, "Classification of mycobacterium tuberculosis in images of ZN-Stained sputum smears," *IEEE Trans. on Information Technology in Biomedicine 14(4),* 2010, pp. 949-957.

[9] C. F. F. Costa Filho, P. C. Levy, C. M. Xavier, M. G. F.Costa, L.B.M. Fujimoto and J. Salem, "Mycobacterium tuberculosis recognition with conventional microscopy," *in 34th Annual International Conf. of the IEEE Eng in Med and Biol Society 1*, 2012, pp. 6263-6268.

[10] C. F. F. Costa Filho and M. G. F. Costa, "Sputum Smear Microscopy for Tuberculosis: Evaluation of Autofocus Functions and Automatic Identification of Tuberculosis Mycobacterium". *In Pere-Joan Cardona (Org.). Understanding Tuberculosis – Global Experiences and innovative Approaches to the Diagnosis: Intech. 1*, 2012, pp. 277-292.

[11] M. Subbarao, and J. K. Tyan, "Selecting the Optimal Focus Measure for Autofocusing and Depth-From-Focus," *in IEEE Trans. on Pattern Analysis and Machine Intelligence . 20(8)*, 1998, pp. 864-870.

[12] M. J. Russel and T. S. Douglas, "Evaluation of autofocus algorithms for tuberculosis microscopy," *in 29th Annual International IEEE EMBS Conference,* 2007, pp. 3489-3492.

[13] C. F. F. Costa Filho, M. G. F. Costa and A. J. Kimura, "Autofocus functions for tuberculosis diagnosis with conventional sputum smear microscopy," *In: Méndez-Vilas editor. Current Microscopy Contributions to Advances in Science and Technology, Badajoz, Spain: Formatex*, 2013, pp. 13-20.

[14] O. A. Osibote, R. Dendere, S. Krishnan and T. S. Douglas, "Automated focusing in bright-field microscopy for tuberculosis detection," *Journal of Microscopy 240(2)*, 2010, pp. 155-163.

[15] M. K Osman, M. Y. Mashor and H. Jaafar, "Performance comparison and thresholding algorithms for tuberculosis bacilli segmentation," *in Proc of International Conf of Computer, Information and Telecommunication Systems 2012*, 2012, pp. 1-5.