

## Image processing techniques for identifying *Mycobacterium tuberculosis* in Ziehl-Neelsen stains

P. Sadaphal,<sup>\*†</sup> J. Rao,<sup>‡</sup> G. W. Comstock,<sup>\*†</sup> M. F. Beg<sup>‡</sup>

<sup>\*</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, <sup>†</sup> University Research Co., LLC, Bethesda, Maryland, USA; <sup>‡</sup> School of Engineering Science, Simon Fraser University, Burnaby, British Columbia, Canada

### SUMMARY

Worldwide, laboratory technicians tediously read sputum smears for tuberculosis (TB) diagnosis. We demonstrate proof of principle of an innovative computational algorithm that successfully recognizes Ziehl-Neelsen (ZN) stained acid-fast bacilli (AFB) in digital images.

Automated, multi-stage, color-based Bayesian segmentation identified possible 'TB objects', removed artifacts by shape comparison and color-labeled objects as 'definite', 'possible' or 'non-TB', bypassing photomicrographic

calibration. Superimposed AFB clusters, extreme stain variation and low depth of field were challenges. Our novel method facilitates electronic diagnosis of TB, permitting wider application in developing countries where fluorescent microscopy is currently inaccessible and unaffordable. We plan refinement and validation in the future.

**KEY WORDS:** tuberculosis; sputum; microscopy; diagnostic imaging; software; Bayes' theorem

SPUTUM SMEAR MICROSCOPY using the ubiquitous Ziehl-Neelsen (ZN) stain, the primary diagnostic strategy for active tuberculosis (TB) worldwide that is recommended by the World Health Organization (WHO),<sup>1</sup> is constrained by its reliance on human skill and time-intensive nature. In developing countries, fewer skilled technicians, inadequate equipment and high caseloads compound the problem of numerous samples at limited microscopy facilities. Reading of three initial smears and viewing of 100 high-power fields per slide is recommended. This takes at least 15 min for experienced staff,<sup>2</sup> is subject to error,<sup>3</sup> and the mental concentration and visual strain limits the volume of slides handled per day. Automated detection of acid-fast bacilli (AFB) could hasten diagnosis, enhance quantitative classification and reduce errors. Two reported studies applying computational image processing to detect TB bacilli used the expensive, less accessible auramine-rhodamine fluorescent stain.<sup>4,5</sup> We demonstrate proof of principle of an innovative computational algorithm to recognize ZN-stained AFB in digital images. Color-edge detection methods applied to fluorescent images work poorly on ZN images. Our new method, free from scale calibration, uses prior knowledge about the distinctive ZN stain color to label a given image pixel as a 'TB object' or a 'non-TB object', followed by shape/size analysis to refine detection.

<sup>†</sup> deceased.

### METHODS

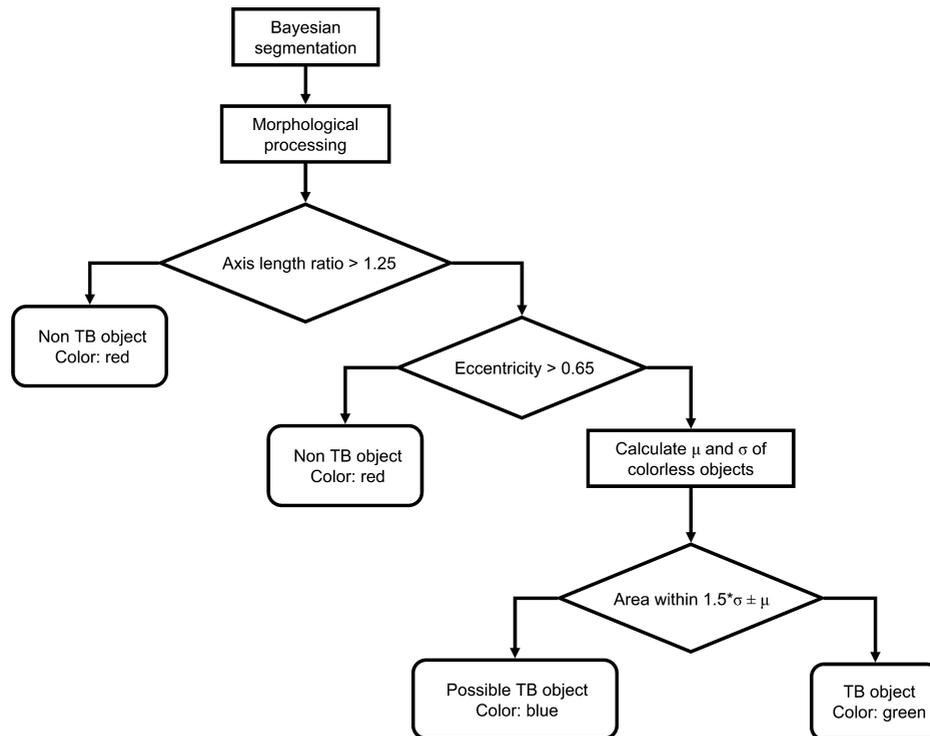
Digital images of ZN-stained smears and tissue were obtained from the Centers for Disease Control's (CDC's) Public Health Image Library (PHIL) laboratories and other sources (Kandavelu K, personal communication, 2006).<sup>6-9</sup> To be selected, images had to depict ZN-stained AFB. For simpler segmentation, we used clear *Mycobacterium tuberculosis* broth images to characterize AFB color. Computational algorithms (Figure 1) were implemented on a personal laptop in MATLAB 7 Image Processing Toolbox software (The Mathworks Inc., Natick, MA, USA, 2004).

#### Color segmentation

Color gradient-based AFB segmentation and edge detection work poorly for ZN stains, which have greater background detail than fluorescent images. We used 1) Bayesian segmentation to predict the probability of a pixel representing a 'TB object' using prior knowledge of ZN stain colors; and 2) shape/size analysis. Pixel probabilities were derived from the three-dimensional probability density function histogram created by manually segmenting ZN-stained AFB-positive images. Most true AFB pixels had red green blue values that were significantly different from non-TB objects. Thresholding pixel probabilities created a binary mask which was improved using morphological dilation of the image with a circular structuring element.

Correspondence to: Mirza Faisal Beg, School of Engineering Science, Applied Sciences Building, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. Tel: (+1) 778 782 5696. Fax: (+1) 778 782 4951. e-mail address: mfbeg@ensc.sfu.ca

Article submitted 18 November 2007. Final version accepted 7 January 2008.



**Figure 1** Flowchart showing the classification steps for automatic identification and labelling of bacilli in the ZN-stained digital images. ZN = Ziehl-Neelsen; TB = tuberculosis.

### Shape extraction

We discriminated true AFB from other AFB-positive artifacts such as *Nocardia* using shape. Matching nearest-neighbor connected pixels were grouped; to account for varying bacillary orientations and magnification, we bypassed size calibration by employing two shape descriptors that were invariant to rotation, translation, skew transformations and scale: 1) axis ratio (1 for circles, higher for line segments) and 2) eccentricity, a ratio of distance between elliptical foci to major axis length (1 for line segments, 0 for circles). The typical axis ratio of 2–2.5 for true AFB was significantly different from approximately ‘one’ for non-AFB objects; similarly, true AFB eccentricity was 0.90–0.96 and centered at zero for non-AFB. To maximize rod-shaped object recognition, we empirically chose conservative threshold cut-offs (axis-ratio  $>1.25$  and eccentricity  $>0.65$ ) as indicating AFB. Objects below the thresholds were labeled red as ‘non-TB’ objects. Calculating the mean AFB size  $\mu$  and standard deviation  $\sigma$  from a broth image, we labeled all size outliers  $> \mu \pm 1.5\sigma$  in blue as ‘possible’ and within  $\mu \pm 1.5\sigma$  in green as ‘definite’ TB objects.

### RESULTS

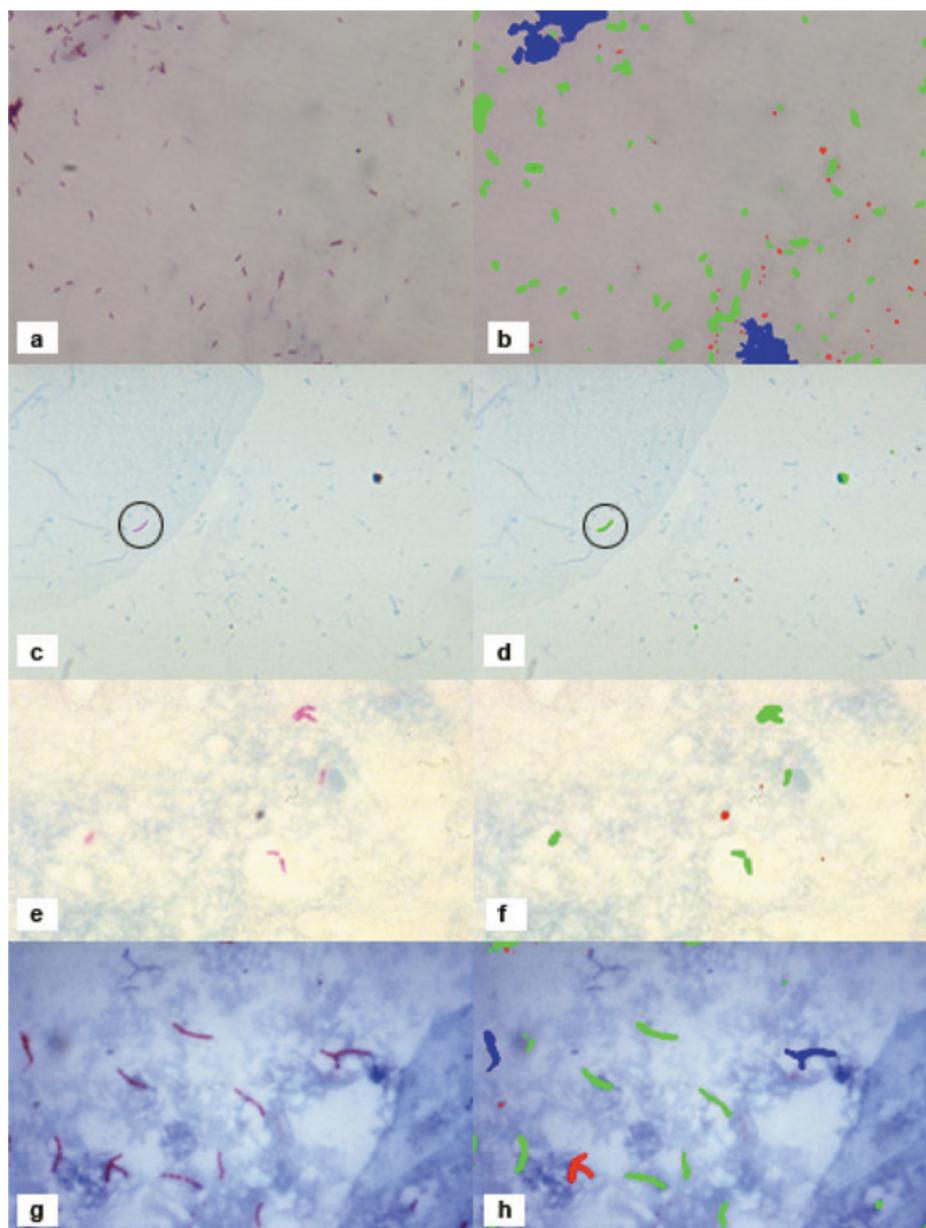
The algorithm recognized AFB under wide latitudes of staining, magnification and resolution (Figure 2). In Figure 2a,b nearly all visible bacilli were color-labeled as TB objects (green); conglomerations were labeled possible objects (blue). In Figure 2c,d the sin-

gle typical TB bacillus was clearly recognized alongside a minor artifact. In Figure 2e,f, all AFB were recognized. In a challenge tissue slide (image not shown), the single TB bacillus was successfully detected without artifacts.

In the typical sputum smear image (Figure 2g,h), the algorithm successfully detected most of the single bacilli; most overlapping bacilli were labeled as ‘possible TB’. In all images, very little of the slide background was incorrectly labeled as possible or definite TB. In an image showing AFB in a skin lesion (image not shown), individual bacilli and small groups were correctly identified, while conglomerations and contiguous background were labeled ‘possible TB’. Even with bacillary conglomerations in AFB-positive images, individual bacilli were likely to be correctly labeled.

### DISCUSSION

We present the first report of automated computational recognition of AFB in ZN-stained digital images, a challenge due to color density similarity between primary stain and counterstain. Our approach has several benefits. First, we specifically targeted the widely used ZN staining method, which is globally relevant to national TB programs. Second, our method succeeds despite considerable background artifacts. Third, our incorporation of size-invariant shape selection allows differing magnification and increases robustness for automated processing without human interaction. The programming simplicity of our algorithm permits con-



**Figure 2** Result of our automated method for labelling TB bacilli. In the left column are the original images and in the right column are the results of the automatic identification. Red objects are those detected by Bayesian segmentation but rejected as TB objects based on shape. Blue objects are those detected by Bayesian segmentation and by shape segmentation but have an incorrect size and are considered as 'possible' TB objects. Green objects are 'definite' TB objects that passed Bayesian segmentation, shape segmentation, and size analysis. TB = tuberculosis. This image can be viewed online in colour at <http://www.ingentaconnect.com/content/iuatld/ijtdl>

version to stand-alone software for use on numerous personal computers in developing countries.

Overstaining or understaining may impact the current algorithm, as Bayesian segmentation uses stored probabilities. However, the algorithm can easily be modified to 'learn' stain characteristics for a new batch of slides by a color calibration step where new sample images are compared to the training database and flagged for differing color histograms. However, staining errors are unlikely in laboratories that follow the quality control procedures recommended by the

WHO.<sup>1</sup> Descriptors for 'Y' or 'T' shaped bacillary conglomerations could be another future improvement. Extensive refinement and validation is planned.

Relevant to centers having few skilled technicians and high workload, our method could reduce TB diagnostic delays. High bacillary load images could be rapidly labeled for immediate analysis and images with few bacilli could be reviewed faster with color-labeling. Our technique could also potentially quantify TB bacilli in ZN smears for future use as a surrogate marker of prognosis or cure in TB patient follow-up and

to evaluate new anti-tuberculosis compounds. Program personnel and researchers could gainfully exploit the convenience and speed of automated processing to accurately and rapidly diagnose active TB worldwide.

#### Acknowledgements

The authors thank R C Ellis, Laboratory Manager, Division of Pathology, Institute of Medical and Veterinary Science, The Queen Elizabeth Hospital, Woodville, South Australia and the Armed Forces Institute of Pathology, Washington DC, USA, for permission to reproduce AFB images from their collections, and K Kandavelu, MBBS, of the Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, for providing digital images of *M. tuberculosis* broth cultures. Thanks are due to N Hooper for suggesting the CDC PHIL database. The authors are grateful to J M Samet, MD, MS, for his thoughtful comments on the manuscript.

This work was supported in part by a grant from the Fogarty TB/AIDS International Training and Research Programs through the Fogarty International Center/US National Institute of Health Grant no. 2 D 43 TW00010-18-AITRP and the National Science and Engineering Research Council grant 31-611387.

Competing interests: The authors declare that they have no competing interests.

#### References

- 1 World Health Organization Global Tuberculosis Programme. Laboratory services in tuberculosis control. Part II: microscopy. WHO/TB/98.258. Geneva, Switzerland: WHO, 1998.
- 2 Central TB Division. Revised National Tuberculosis Control Programme: module for laboratory technicians. 2nd ed. New Delhi, India: Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India, 2005.
- 3 Nguyen T N, Wells C D, Binkin N J, Pham D L, Nguyen V C. The importance of quality control of sputum smear microscopy: the effect of reading errors on treatment decisions and outcomes. *Int J Tuberc Lung Dis* 1999; 3: 483–487.
- 4 Forero M G, Cristobal G, Alvarez-Borrego J. Automatic identification techniques of tuberculosis bacteria. *Proceedings of SPIE* 2003 Nov 1; 5203 Applications of Digital Image Processing XXVI, 2003: 71–81.
- 5 Veropoulos K, Learmonth G, Campbell C, Knight B, Simpson J. Automated identification of tubercle bacilli in sputum. A preliminary investigation. *Anal Quant Cytol Histol* 1999; 21: 277–282.
- 6 Ellis R C. Gallery of histological sections: acid-fast bacilli. Woodwill, South Australia, Australia: IMVS Division of Pathology, The Queen Elizabeth Hospital, 2004. <http://www.adam.com.au/royellis/zn.html> Accessed January 2007.
- 7 Centers for Disease Control and Prevention. Public health image library. Atlanta, GA, USA: CDC, 2007. <http://phil.cdc.gov/phil/home.asp> Accessed January 2007.
- 8 Armed Forces Institute of Pathology. Department of Veterinary Pathology. Washington DC, USA: AFIP, <http://www.afip.org/vetpath> Accessed January 2007.
- 9 Raqib R. International Centre for Diarrhoeal Disease Research, Bangladesh. Dhaka, Bangladesh: ICCDDR, B. <http://www.icddr.org> Accessed March 2007.

#### R É S U M É

Dans le monde entier, les techniciens de laboratoire s'ennuient à examiner les frottis de crachats pour le diagnostic de la tuberculose (TB). Nous avons démontré par preuve-de-principe un algorithme computé novateur qui reconnaît avec succès des bacilles acido-résistants (BAAR) colorés au Ziehl-Neelsen dans les images digitales.

Une segmentation Bayésienne automatisée à étapes multiples et basée sur la coloration a identifié des éléments possiblement TB, a retiré des artefacts par comparaison des formes et a identifié par les couleurs des éléments comme « certainement », « peut-être » ou « non-

tuberculeux » en court-circuitant la calibration photomicrographique. Les défis rencontrés ont été les amas superposés de BAAR, des variations extrêmes de la coloration et une faible profondeur de champ. Notre méthode novatrice facilite le diagnostic électronique de la TB, en permettant une application plus large dans les pays en développement où la microscopie par fluorescence est actuellement non-disponible et trop coûteuse. Nous envisageons pour l'avenir un raffinement et une validation de cet algorithme.

#### R E S U M E N

En todo el mundo, los técnicos de laboratorio realizan tediosamente la lectura de las baciloscopias de esputo para el diagnóstico de la tuberculosis (TB). El presente artículo constituye el estudio demostrativo preliminar de un algoritmo informático que reconoce exitosamente los bacilos acidorresistentes con la tinción de Ziehl-Neelsen en imágenes numéricas.

Mediante una segmentación bayesiana de múltiples fases basada en el color, se identificaron posibles 'objetos tuberculosos', se descartaron artefactos por comparación de formas y se marcaron los objetos con colores como

'seguros', 'posibles' o 'no tuberculosos', eludiendo la calibración de las fotomicrografías. Los agregados de bacilos superpuestos, la variación extrema de la coloración y la escasa profundidad de campo constituyeron obstáculos al procedimiento. Este nuevo método facilita el diagnóstico electrónico de la TB y permite una aplicación más amplia en los países en desarrollo donde no se tiene acceso a la microscopía de fluorescencia o es inabordable. Los autores proyectan afinar y validar la técnica en el futuro.