Tuberculosis: Image Segmentation Approach Using OpenCV

Abdullah Ayub Khan*   Anil Kumar†   Gauhar Ali†

Abstract

Tuberculosis (TB) is one of the major disease spreading all over the world. TB caused by bacteria is known as Mycobacterium tuberculosis. Nowadays, TB is increasing widely in the region of Karachi and now it's becoming a challenging task for all researchers. The process is to partition the digital image into different segments according to the set of pixels known as image segmentation. It's used to find segments & extract meaningful information of an image. Image segmentation approaches are providing new ways in the field of medical and it's exactly suitable for TB images, block-based & layer-based segmentation helps to identify edges, thresholding, regional growth, clustering, water shading, erosion & dilation, utilizing histogram for the betterment of TB patients. Chest X-ray is playing a vital role to diagnose TB rapidly. TB image contains binary colors, it's either black & white but it would have been a different level of the color shades. Diagnosing symptoms and intensity of TB in a patients' x-ray is such a critical problem. The purposed solution is to overcome the problem and reduce the ratio of TB patients in Karachi region by using image segmentation approaches on chest X-ray and calculates the alternative way to detect the intensity level of TB in individual patient's report with effectively, efficiently & accurately with a minimum amount of time by using Python OpenCV.

Keywords: Image Segmentation Approaches, Tuberculosis (TB), Medical Imaging, Binary Color, Python, OpenCV

1.   Introduction

Tuberculosis (TB) is becoming a hardly manageable disease in recent era throughout the world, the rate increasingly goes up in the region of Karachi. TB caused by a bacterium named Mycobacterium Tuberculosis (MTB). It mostly effects on lungs but sometimes it infects on other organs in the human body. It can spread from one person to another through the air. The first TB infection happened about 9,000 years ago [1]. According to researches, it's the second biggest killer disease in the world. In 2015, 1.8 million people died and 10.4 million people fell ill by tuberculosis [2]. The ratio is going to increases day-to-day. In the list of world populations, Karachi is in 3rd position[1] [3]. Last few years, TB expand rapidly in the region of Karachi. According to "National TB Control Program", every year TB kills 90000 people in Pakistan. In Karachi 2010-2013, more than 14000 TB patients were registered [4]. In 2016, Sindh 4th Quarter Tuberculosis (FTI) survey shows 15290 patients infected by all types of TB and 6798 patients are newly registered [5], that's meant more than 22000 patients are affected by TB in Sindh. In short, tuberculosis killed more "Karachiites" than terrorist did.

A single picture translates more information about the scene than a human can. Image segmentation (IS), the process in which an image is converted into multiple portions. These portions are used to find objects, features or related information of a digital image. The objective of IS is to analyze coherent objects, pixels, color, shapes, corner, edges, etc. for the meaningful understanding of an image. IS is also used to label each pixel and these labeled pixels has some specific characteristics. There are some certain techniques of an Image Segmentation which are: thresholding, region-based segmentation, regional growth, edge detection, filtering, hybrid segmentation (water shading). Such techniques are used in chest x-ray of Karachi's patients to identify TB segmentation using Python OpenCV.

The proposed solution tries to minimize time computation as-well-as reduce cost and increase productive

---

*Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan
†Sindh Madressatul Islam University, Karachi
Corresponding Email: abdullah.khan00763@gmail.com
[1]http://www.citymayors.com/statistics/largest-cities-population-125.html

treatment of TB patients using x-ray-based image segmentation. By the help of this, TB can easily be diagnosed, treatment can easily be started without waiting for other report. Chest x-ray image segmentation techniques can be applied for finding the intensity (color, shape, texture based, etc. using thresholding and edge detection etc.) of TB. In this paper, the proposed solution will answer these queries: which part of the lungs is affected? what category of TB patient has? it affects lungs first time or not? how to prevent? how it will take time to overcome this problem? how to apply IS techniques? how computer vision helps to find the solution in medical imaging? and many more. The motive is just scan patient chest (x-ray), diagnose the symptoms of TB, ensure the category of TB, need to know the way of treatment, start treating without wasting of time. It's like a report less treatment.

2. Related Work

Segmentation of organs accurately using chest x-ray is a well-defined problem in the field of medical imaging, find coherent objects of an image and extract useful knowledge in it which help to move one step ahead in medical field. Several papers published related to IS and TB (chest x-ray) in past few years. Some latest literature reviewed in this section, crucial key factors are discussed below:

Nida M. Zaitoun et.al briefly elaborates the importance of image segmentation (IS) in image processing (IP). IS is not just finding edges (coherent object) of an image, but there is a lot of other feature which help identify complete sense of an image. Methods for IS splits into two parts block-based segmentation & layer-based segmentation. In block-based segmentation, divide into two main categories region based & edge based. Region based methods contains Region Growth, Split & Merge, Clustering, Thresholding and Normalized cut. Same as, edge-based methods contain Roberts, Sobel, Prewitt, Canny. Soft computing approaches has famous algorithms like Neural Networks, Genetic Algorithm, Fuzzy logic [6]. N. Dhanachandra et.al highlights some crucial factors of IS IS is the first step of IP. Article explains the importance of clustering algorithms in IS. IS contains lots of techniques but clustering provides advance features in the field of IP. Although it is derived by block-based segmentation, the objective of clustering is to classify clusters of different objects. To categorize clusters, we need algorithms like k-mean, fuzzy c-mean, subtractive, expectation and maximization, DBSCAN. Each algorithm has different ability to group data (cluster) of an object in an image [7]. Pixel is a prime factor in medical imaging. Group of pixels of different objects can be used to observe similar data and easily locate the point of interest of an image where we can easily analyze.

Frieze, Julia B et.al defines the ratio of childhood TB in Cambodia. The childhood TB cases are increasing rapidly with $10\% - 20\%$ of total TB cases. In between 2015-2016, diagnoses half a million new cases emerged and almost 74000 people died annually because of TB. In adults, it can easily diagnose TB for analyzing chest x-ray reports but in children, most of the time it can't be detected because it's difficult to diagnose as-well-as hospital needs latest equipment & technology in Cambodian hospitals. Diagnoses of TB in children is such a challenging task. It exceeds 87% in the last year. The proposed solution, Cambodian's provinces are divided into Operational Districts which cover 1,80,000 people. These Operational Districts report to National TB Program in Cambodia. Overcome of childhood TB diagnosis, take an interview of each child parents or guardian who have been suffering from TB since childhood. After gathering data, deliver knowledge to hospitals and provides training [8]. TB isn't the problem of Pakistan & Cambodia, but it spread all over the world. World should take some necessary action to eliminate this killer disease assoonas possible. Rangaka, Molebogeng X et al suggested the idea related to reduction of tuberculosis infection in the globe.The roadmap provided by researchers with implementation barriers and challenges. The solution based on the clinical and technical approach, health-systems, policy & leadership, advocacy approach [9].

Color image segmentation of tuberculosis bacilli in ziehl-neelsen-stained tissue image using clustering approach. In clustering, moving k-mean algorithm can segments group of TB infection manipulate into color-based. The original image which is based on RGB can be converted into C-Y transformation, applying k-mean algorithm with median filters for removing noise, after that regional growing can separate image into multiple regions, finally image can be segmented properly for the detection of TB bacilli [11]. Raof, M. Y. Mashor, and S. S. M. Noor segmented TB bacilli in ziehl-neelsen sputum slide images using clustering algorithm. Separating foreground & background of medical image with accuracy plus efficiency. Modification of medical imaging, segmentation is performing a vital role [12]. The idea of automated image segmentation proposed by Riza et.al. The step-by-step method interprets the overall scenario of automated image segmentation of TB bacilli. It starts with image contrast which enhances image in order to clear and brighten, after contrasted color space can be done for detecting infections, change RGB color into image label and image clustering pixels image into multiple color objects, at last it can be segmented exactly [14]. Image processing is also effective for diagnosing TB bacilli. This research is done on MATLAB software tool for detecting and counting TB bacilli using color-based approach segmentation with accuracy [13]. Machine learning and knowledge-based system also contributes in the medical imaging (MI) field. Melendez, Jaime et al describe the computer aided detection using supervised learning & deep learning for MIS [10].

The above literatures having reviewed, we come to the point that majorly IS working can be done with machine learning or clustering algorithm, to group similar data or infection. Most of the time it can be done on MATLAB or other well-known popular tools (as already mentioned above). So, we are utilizing the efficiency of Python interact with OpenCV for extracting the highest percentage rate of accuracy of TB images with minimal amount of time.

3. Proposed Methodology

The proposed solution elaborates unique identification of TB in short period of time using OpenCV 3.4, Matplotlib & NumPy collaborate with Python 3.6.5. We explain the importance of image segmentation in the field of medical.
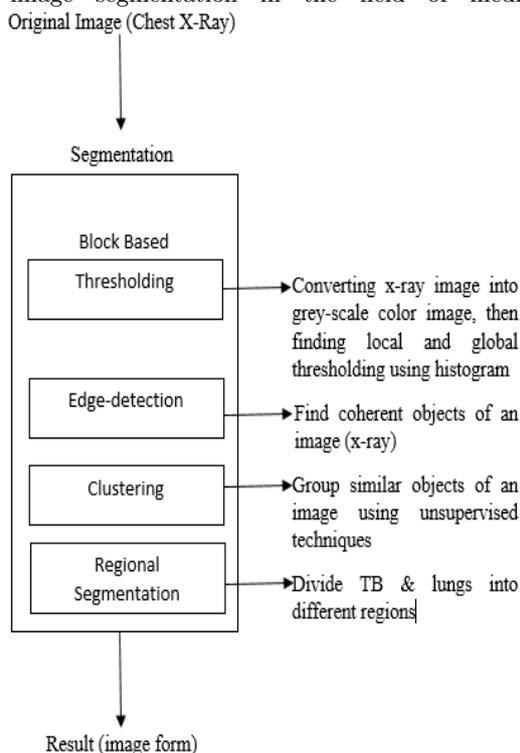

Figure 1: Graphical Representation of The Proposed Solution

In this research article, mentioned above the graphical representation of proposed solution can appropriate for diagnosing TB intensity with minimal amount of time. In the first step (pre-processing), applying filtration of an image for removing noise and enhancing image quality in terms of smoothing, sharpening, and restoring. In the next step, we perform several activities of block-based segmentation, convert image into grayscale histogram and applying thresholding to find

the impact of TB on lungs. After, separation the coherent object on lungs by using edge detection. Cluster means group similar objects, k-mean algorithm utilizes for finding k-neighboring. At the end, an image can be separated into two regions, one is infected by TB and the other is the lungs. Below the result section shows the overall mechanism of segmentation approaches with detailed description and graphical view.

4. Tools & Packages

In this context, considering crucial parts of the research is to pick suitable tool, packages and programming language. These things play a vital role in our research for analyzing image data & retrieving meaningful information. The next two sections define the importance of tools for managing and maintaining tuberculosis patients' data and transform it into useful manner.

4.1 Python

Python is one of the powerful tools for making program, projects and portfolio. Program in terms of creating different projects for performing specific task and it can reduce load of the machine. Python is a programming paradigm which supports lots of programming abilities like object-oriented, structural, high-level, functional, interpretation and dynamic scripting skills[2] . There are two main versions named: latest version (it starts with 3.0 or so on) and popular version (2.7 or so on). In this research, we follow the latest version which is python 3.6.5, it gives complete programming facilities including built-in functions like: list-comprehension, slicing, dictionary, corpus, lambda, set, sort, min/max, reverse, user define function (UDF), etc. which reduce the programming complexity; by this act it increases efficiency. Python provides functions which decreases line of codes and increases accuracy. In python, lots of external open source libraries available on internet some most important: nltk (nltk corpus, tokenization, stemming), Pycrypto, OpenCV, Matplotlib, NumPy & many more. There is no need to learn each and everything, each library is expressive in nature meant that all is similar with each other. Just extract it and use it as per need. Providing functionality and simplicity of context, user can easily understand as use it without any difficulty. In this nature, we utilize this powerful language in our research with interact other library for segmentation.

4.2 OpenCV

In this article, our main focus is on OpenCV just because of image segmentation. OpenCV is an open source library which is used in computer vision field conjointly with python to make a useful program for specific task. More than 2500 optimization algorithms including comprehensive of both state of the art and

---

[2]https://www.python.org/doc/essays/blurb/

classic computer vision as-well-as machine learning algorithm. Utilization of all these algorithms we get line detection, edge detection, corner, moving objects, motion sequence, feature detection and recognition, segmentation, image stitching, field of view (panorama), human motion detection, human gesture & posture detection of suspicious person, face detection, finger-print detection and much more. Number of downloads exceeded more than 14 million till now. Is has an ability to collaborate with C++, java, python, MATLAB and it provides interface for all operating systems like Mac OS, Linux, Window, Android. There is an extensive use of OpenCV through all over the Globe. Many big companies utilizing OpenCV for completing so many CV (Computer vision) tasks. OpenCV is written in C++, it's easy to use and provides lots of functionalities for design and implementation of CV products.

5. Dataset

Karachi X-Ray[3] shared with us the crucial data of Karachiites' TB patients. The data is based on images form (chest x-ray) which were collected at the beginning quarter of the year 2018 (shown in figure 2). In this article, there are thirty-four different images of different patients including men, women & children. These people live in Karachi. New patients are also recorded in the year 2018. Many lives survive against world's challenging problem for many years. But it never is decreased since recent year, although it spread one to another person rapidly. In fact, the contagion is also in a newly born baby as well. Apply image segmentation on it and overcome with a different solution in computer vision field.
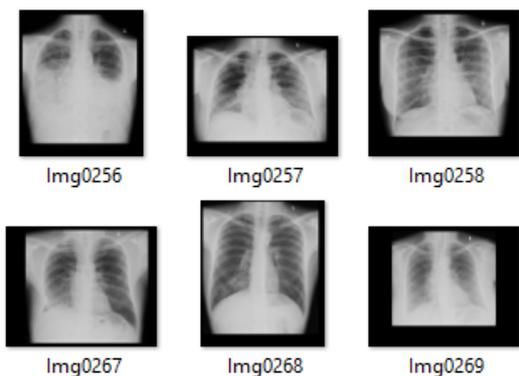

Figure 2: TB Original Images

6. Result & Discussion

In result section, we are describing the overall mechanism using in this research, detection of tuberculosis in lungs x-ray images. There are important steps which analyze image, extract information, retrieve & store.

[3]The number one health diagnostic center (link: http://karachixrays.com)

Block-based image segmentation approaches segments coherent image in various manner like thresholding, edge detection, filtration (low pass & high pass), regional growth, and clustering. The steps are mentioned below with some description & graphical representation.

Filtration:
The first step is preprocessing, removing noise. Filtering is the process which enhances the image features. There are two main types of filtering, low-pass and high-pass filter. In this research, kernel convolution 3x3 n-d matrix is used to smoothing, edge enhancement and sharpening of the images (shows in figure 3).

```
[[-1. -1. -1.]
 [-1.  8. -1.]
 [-1. -1. -1.]]
<class 'numpy.ndarray'>
```
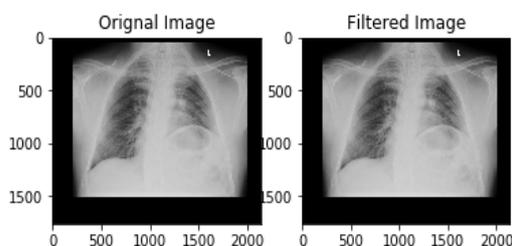

Figure 3: Removing Noise Using 3x3 Kernel n-d matrix

Thresholding:
It converts grayscale image into binary color (0 & 1) and find intensity of black & white portion of the image, lungs can be detecting as black color and white shows how much TB affected on lungs (shows in figure 4). There are several thresholding approaches available, the research adopts five different forms of thresholding, which are: 'Binary', 'Binary Inverted', 'Zero', 'Zero Inverted' & 'Trunc' (shows in figure 5. In this research, we set thresholding value between 127 to 255, for clear understanding of binary image intensity (see some graphical representation in histogram).
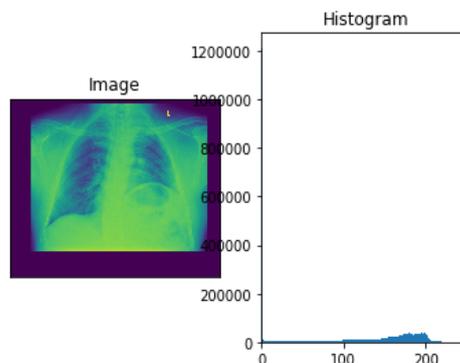

Figure 4: Thresholding

It's an image processing technique used to detect boundaries of objects in an image. In edge detection, various types of algorithms already developed, canny edge detection is appropriate for medical imaging. Gaussian kernel 5x5 n-d matrix, intensity gradian 'L1' and 'L2' norm, L1 level of intensity sets between 20 to 40, and L2 between 20-30 ratios (shows in figure 6).
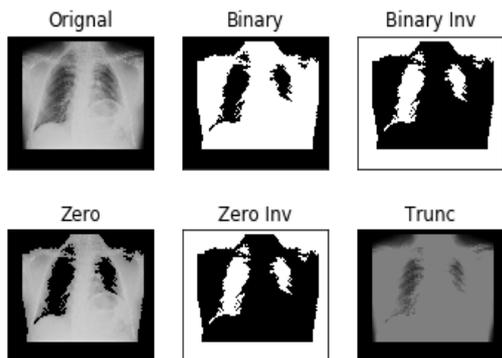

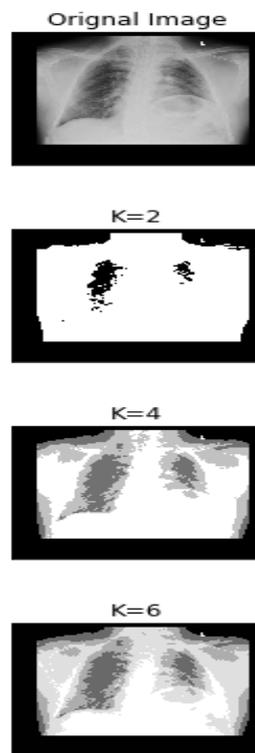Figure 5: Types of Different Thresholding Applied

Figure 6: Canny Edge Detection

Clustering:

In clustering, segmentation can be done on group of similar objects (cluster) in an image. Unsupervised, no labelling, K-mean clustering algorithm used to assemble similar objects. In the research, we set kth value as k=2, 4, and 6.

Figure 7: Clustering (K-mean with different K points)

Regional-Based Segmentation:

Separate an image background and foreground into different regions. Furthermore, segmentation of an image objects, color, shapes, texture, and more features are becoming different regions (figure 8 & 9). Choosing the interested region for segmenting and extracting hidden pattern or meaningful knowledge in that image.
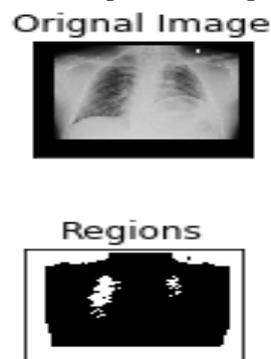

Figure 8: Clustering (Separating Regions of an Image)

Erosion

Dilation

Figure 9: Clustering (Erosion & Dilation of an Image)

7.   Conclusion

Medical imaging is the hot topic nowadays, detecting diseases in a human body is a critical task for all researchers. Tuberculosis (TB) is the emerging problem in the Karachi region, resolving such type of situation we need some tools and techniques. Computer vision provides a different path for recognizing objects in an image, and machine learning algorithms help identify efficiently. Image segmentation approaches recognize image into different segments such as color, texture, shapes, size, edge, objects, and regions. Categories segmentation into three main parts but block-based segmentation is suitable for detecting TB in x-ray. In block-based, segmentation can be done by filtering, thresholding, edge detection, clustering & regional growth. In this research, we applied all those techniques and elaborate on the importance of each. OpenCV, NumPy and Matplotlib try to summaries code of thresholding, Canny edge detection, clustering, and regional growth in python 3.6.5, given image as an input and display image as an output but get hidden patterns or meaningful information. The result section clearly shows, block-based image segmentation is one of the best solutions for medical imaging, it separates accurately background and foreground of an image (chest x-ray), which help to detect the intensity of TB on lungs with a minimal amount of time. Our main objective, to ensure a better understanding of IS approaches in the medical field which help diagnose diseases, take suitable action in terms of treatment and reduce the rate of patients recodes in health sector department.

8.   Future Work

Tuberculosis is widely expanding not only in Karachi but overall in Pakistan. Millions of people affected and new cases emerging in regular bases. For successfully done image segmentation approaches in the region of Karachi. Now, our next target to apply the same scenario in whole TB patients of Pakistan.

References

[1] Hershkovitz, Israel et.al (2008-10-15), "Detection and Molecular Characterization of 9000-Year-Old Mycobacterium tuberculosis from a Neolithic Settlement in the Eastern Mediterranean".

[2] McIntosh, James. "All you need to know about tuberculosis." Medical News Today. MediLexicon, Intl., 27 Nov. 2017.

[3] City Mayors Statistic March 2018, "Largest Cities in the World (1-150)", link: http://www.citymayors.com/statistics/largest-cities-population-125.html

[4] Miandad, Muhammad & Burke, Farkhunda & Nawaz-Ul-Huda, Syed & Azam, Muhammad. (2014). Tuberculosis incidence in Karachi: A spatio-temporal analysis. GEOGRAFIA, Malaysian Journal of Society and Space. 10. 01-08.

[5] Pakistan Bureau of Statistics, "T.B Report_08_06_2017".

[6] Zaitoun, Nida M., and Musbah J. Aqel. "Survey on image segmentation techniques." Procedia Computer Science 65 (2015): 797-806.

[7] Dhanachandra, Nameirakpam, and Yambem Jina Chanu. "A survey on image segmentation methods using clustering techniques." European Journal of Engineering Research and Science 2.1 (2017): 15-20.

[8] Frieze, Julia B., et al. "Examining the quality of childhood tuberculosis diagnosis in Cambodia: a cross-sectional study." BMC public health 17.1 (2017): 232.

[9] Rangaka, Molebogeng X., et al. "Controlling the seedbeds of tuberculosis: diagnosis and treatment of tuberculosis infection." The Lancet 386.10010 (2015): 2344-2353.

[10] Melendez, Jaime, et al. "A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays." IEEE transactions on medical imaging 34.1 (2015): 179-192.

[11] Osman, M. K., et al. "Colour image segmentation of tuberculosis bacilli in Ziehl-Neelsen-stained tissue images using moving k-mean clustering procedure. " Mathematical /Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on. IEEE, 2010.

[12] Raof, Rafikha, M. Y. Mashor, and S. S. M. Noor. "Segmentation of TB Bacilli in Ziehl-Neelsen Sputum Slide Images using k-means Clustering Technique." CSRID (Computer Science Research and Its Development Journal)9.2 (2017): 63-72.

[13] Payasi, Yoges, and Savitanandan Patidar. "Diagnosis and counting of tuberculosis bacilli using digital image processing." Information, Communication, Instrumentation and Control (ICICIC), 2017 International Conference on. IEEE, 2017.

[14] Riza, Bob Subhan, et al. "Automated segmentation procedure for Ziehl-Neelsen stained tissue slide images." Cyber and IT Service Management (CITSM), 2017 5th International Conference on. IEEE, 2017.