CANCER COLONY IDENTIFICATION VIA DIGITAL IMAGE PROCESSING

Barry Dunn, Gavin Mantica, Jessica Jensen, Joshua Kruzan, Juliette Caffrey
Mentor: Ahmet E. Akosman

School of Engineering, Computing and Construction Management, Roger Williams University, Bristol RI

Introduction

In this study, a standardized method of quantification for researchers to develop treatments for lung cancer, which is the most lethal form of cancer, is presented. To administer the proposed tool, a numerical analysis method was created for preclinical use, in which it will accurately and affordably quantify the presence of cancer colonies on mouse lungs in an automated fashion. The quantification algorithms are developed for two different image types: whole lung images and tissue cross section images taken of microscope slides which can be seen in Figure 1.
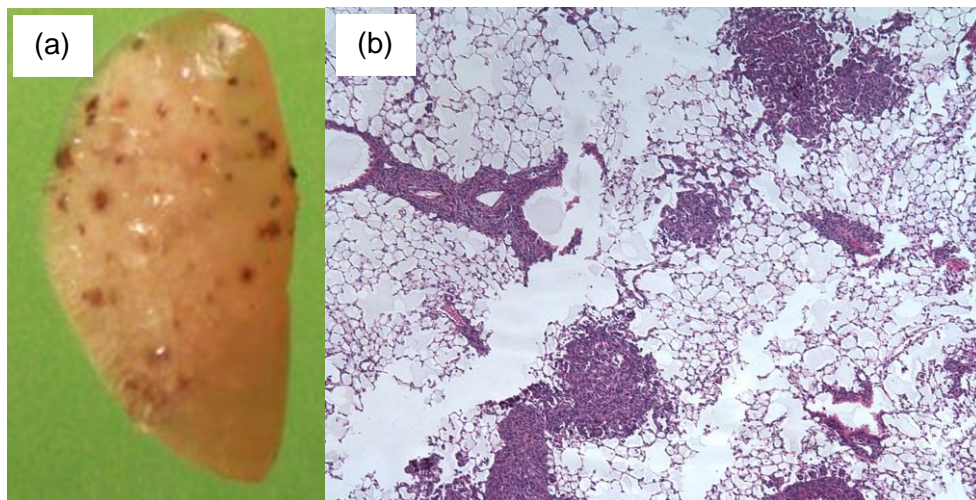

Fig.1 (a) Image of a whole mouse lung and (b) image of a cross-sectional tissue of a mouse lung. In both images, the dark spots, which are the cancer cell colonies, are clearly visible due to the natural pigments of the cells in whole lungs and due to dying the sample in the cross-section image, respectively.

In Figure 1, there are inherent differences between the two types of images, therefore in order to properly analyze the images, it is necessary for them to undergo different forms of digital image processing. To handle this, two different algorithms were developed, to examine whole lungs and their cross sections separately.

Results

For cross-sectional tissue images, it is important to observe that there is both healthy and cancerous tissue on the slide. Also, the less dense matrix of healthy cells should not be included with the dark dense regions in the measuring of cancer colonies. Thus, an adaptive filtering scheme is applied to the gray-scaled image of the original sample to remove any potential shadows before the operation of binarization. In order to outline the cancerous cell clusters, an edge detection method is developed.

Further, size-restricted feature detection algorithms are proposed to avoid remaining miniscule clusters of the healthy tissue is the image. In the last step, a parent/child object detection algorithm is developed so that the empty white space inside the clusters was not counted, allowing for a very precise area calculation. The area was calculated as the percentage of the image by counting the outline pixels versus the total number of pixels in the image. The results of this analysis compared to manually drawn areas is shown in Figure 2.
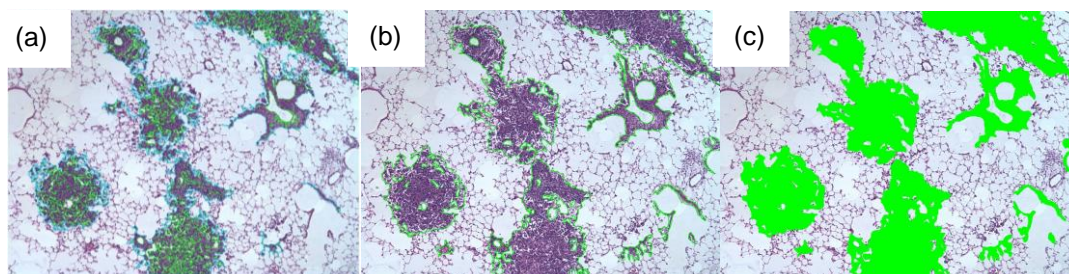

Fig. 2 (a) Lung cross section with the cancer colonies detected and highlighted from the numerical analysis, (b) cross section image with the cancer colonies highlighted manually, (c) the colonies shaded in manually for the area calculation. Agreement between the results in (a) and desired results in (b) is improved as image (a) does a better job excluding the white areas then by hand in image (c).

In Figure 2, image (a) and image (b) have the desired agreement as to what areas should be counted as cancer cell colonies. It can also be seen that the colonies shaded in manually in image (c) include a lot more undesired white space than in image (a). The numerical method appears to provide more accurate results than calculating by hand, and error quantification is currently under development.

The second part of the numerical analysis includes the cancer colony identification in whole mouse lungs. The main characteristics that distinguished the whole lung images from the cross-section images are quite different. First, the lungs do not necessarily fill the entire image, which indicates that the outline of the whole lung needed to be specifically identified in addition to the cancer colonies requiring the image to be filtered in two independent processes. The whole lung outline was detected by converting the image to an HSV domain for looking at the hue channel. This method of filtering was effective due to the second main characteristic of the image being that it was against a green background. To find the outline of the cancer colony clusters, the original image is converted to grayscale and contrast adjustment techniques are applied. After the application of selective filtering, cancer colony clusters are identified using an adaptive threshold technique, which required the need of the gray-scaled version of the original image. Figure 3 shows the results of both the whole lung outline and the cancer cell colony outlines.
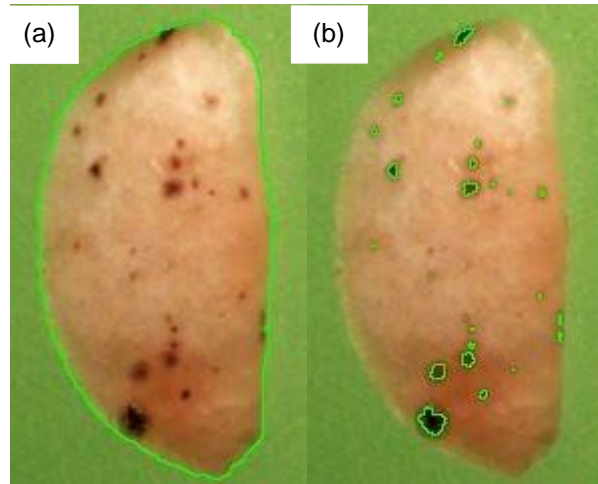
Fig. 3 (a) whole lung image outlined using HSV (b) the cancer cell colonies outlined using adaptive threshold. All the detected shapes are fully enclosed in order to measure the area.

As seen in Figure 3, the outlines for the lung boundary and cell colonies have a great agreement within complete closed shapes. This indicated that the results can be used to effectively calculate the area for both objects through the binarized version of the original image. To determine the parameters in preprocessing methods for tumor detection, a parameter sweep for the edge thresholding strength was executed while varying the parameter of interest. The results were plotted to allow for simple analysis in the determination for the value of the parameter. Figure 4 shows the results of this process for the case of local contrasting.
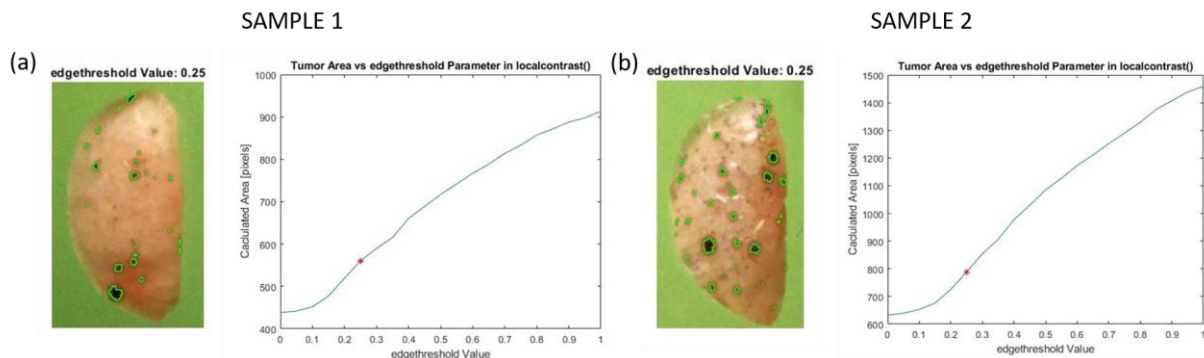

Fig. 4 Detected tumor area on the image of Sample 1 (a) and Sample 2 (b) with the calculated tumor area through the edge thresholding strength parameter sweep.

At the chosen value, indicated in red on each of the plots, it was found that at this point, the area begins to increase linearly with an increase in the parameter. The optimal edge thresholding strength was found for the values higher to the initial plateau of the detected areas for lower values of the strength. By manual inspection, most accurate results are achieved with a strength of 0.25 which is in the beginning of the linearly increasing curve, indicating that the developed algorithm starts detect based on background darkness level instead of the actual edges. This parameter sweep test allows to find the optimal edge thresholding strength without detecting artificial features solely based on pixel values.

An important part of the process in developing a new numerical method is the validation. To achieve this goal, percent error will be quantified for both image types. This will be done by having a large sample size of experts outline and shade in the cancer clusters for each image type. The area shaded will then be calculated for each image and used to find a distribution of the area value based on human approximation. The numerical method will then be used to calculate the area on these images, and statistics will be used to see whether the values fall within the population to a statistically significant degree. This process will then be repeated on multiple images of each type to show that the numerical method is accurate for a variety of images of the same type.

Conclusions

The current standard of practice for the quantification of the presence of cancer cell colonies is to count the number of visible tumors by eye and include an image of the sample; thus, there is no current method for measuring the area of the tumors. Therefore, this method will offer a significant improvement for both speed and accuracy. The numerical method aims to run faster and more consistently than the current standard. The final product is going to be designed for ease of use to researchers. For this reason, the numerical methods are written as MATLAB scripts that can also be run using OCTAVE, meaning that it can be run using free software. A user's manual will be made available; however, the only user input required will be the image of a cancerous sample for which the analysis is desired. There will also be a version of the tissue sample numerical method which will be capable of running off a live microscope feed; however, this version will require access to a specialized MATLAB toolbox.

Acknowledgments