An Assessment of Thresholding, Textural Features and Edge Detection in Fatty Liver Segmentation

Abdullah Al-Hayali 0969687
School of Engineering
University of Guelph
Guelph, Ontario
aalhayal@uoguelph.ca

Miriam Naim Ibrahim 0922788
School of Engineering
University of Guelph
Guelph, Ontario
mnaimibr@uoguelph.ca

***Abstract*—This paper presents prelimary techniques for automated liver segmentation using intensity thresholding, texture features, and edge detection algorithms. Non-alcoholic fatty liver disease (NAFLD) is the most predominant form of chronic liver disease in the western world. NAFLD can be diagnosed through an expert radiologists analysis of ultrasound images. Liver segmentation in ultrasound images is a necessary first step for automated liver disease detection, allowing for more accessible diagnosis. 10 liver ultrasound images were used for testing, and a maximum average accuracy of 79.43% was achieved.**

***Index Terms*—Segmentation, Liver, Thresholding, Edge Detection, NAFLD, Ultrasound, Entropy, Noise, Autocorrelation**

I. Introduction

Non-alcoholic fatty liver disease (NAFLD) is a liver condition that affects populations consuming little to no alcohol [1]. With about one quarter of the United States' population being diagnosed with NAFLD, it is classified as the most common form of chronic liver disease[1]. Although not fully understood, NAFLD can be caused by a range of health conditions such as obesity, high blood sugar, and high blood fat levels. All the aforementioned potential causes are very prevalent in the western world due to the lifestyle and diet [1]. NAFLD, as the name suggests builds fat in the liver tissue, along with fluid build up, swelling of the veins, and potentially liver cancer or cirrhosis (scarring) in more severe cases, therefore a diagnostic method that is accurate will aid clinicians in better and potentially early identification, and treatment[2].

Several medical imaging modalities are available to detect the body's tissues and report findings. Ultrasonography is deemed to be one of the most accurate and efficient modalities in detecting early to severe cases of fatty liver, with accuracies similar to those of histology, which is the current golden standard for detection [2]. The advancements in image analysis techniques provide an opportunity for relatively accurate, automated detection. This can be achieved through a combination of several algorithms such as segmentation, textural classification, and edge detection.

Segmentation is an image analysis technique in which image pixels are clustered in salient image regions, where each region corresponds to individual surfaces, objects, or natural parts of objects [3]. Ultrasound image segmentation delineates the boundaries of structures to automate dimensions, or characterize tissue regions. Ultrasound tissue characterization is an approach used post segmentation to model and analyse ultrasound signals to differentiate between healthy and diseased tissues[4].

In this paper, several image segmentation techniques including edge detection and thresholding will be investigated, with the aim of obtaining high accuracy liver segments from healthy and fatty liver ultrasound images.

II. Methods

A. Data Pre-Processing

The data-set selected was retrieved from Kaggle, and consists of 10 B-Mode abdominal ultrasound images [5]. Diagnosed liver disease, based on biopsy results were also provided. 3 patients are known to be healthy, and the remaining 7 patients are diagnosed with fatty liver disease, having greater than 5% fatty hepatocytes with steatosis. The raw images were provided in a .mat file. A MatLab script was written to produce .png images which were then segmented in Python. The raw images contain undesirable annotations, text, grayscale bars, and graphs as seen in Figure 1.

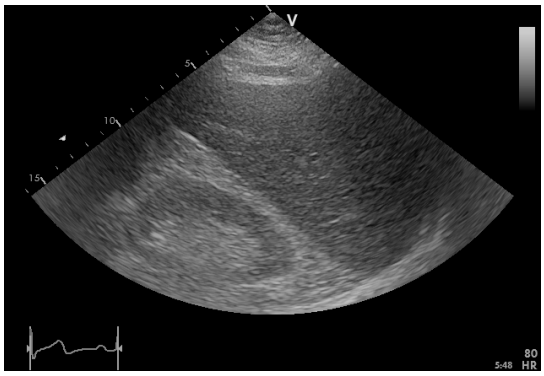

Fig. 1. Raw Ultrasound Image [5]

Thus, a cone-shaped mask was produced to crop the true ultrasound data and eliminate all undesirable features in the image. Furthermore, Kaggle does not provide ground truth liver segments. Masks for the 10 images were manually

created by a radiologist-trained individual, who is experienced in liver ultrasound analysis and quality control. An example of a cropped image and ground truth mask can be seen in Figure 2 below.

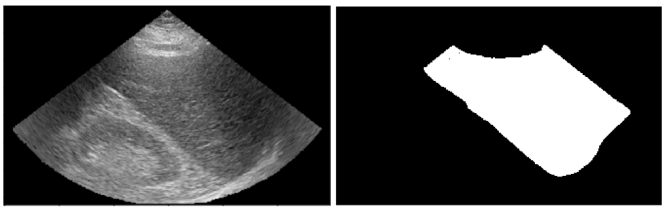

Fig. 2. Example Cropped Ultrasound and Ground-Truth Liver Segmentation Mask

B. Accuracy Measures

As progressive attempts were made to segment the liver, success was determined based on subjective image analysis by the authors, as well as quantitative measures of accuracy. Using the Ground-Truth Liver Mask, and comparing them to the results of various segmentation methods, two forms of accuracy were calculated:

1) Segmentation Accuracy: $\frac{L_t+NL_t}{L_t+NL_t+L_f+NL_f}$
2) Liver Identification Accuracy: $\frac{L_t}{L_t+L_f}$

Where:

- L_t = Correctly classified liver pixels
- NL_t = Correctly classified non-liver pixels
- L_f = Incorrectly classified liver pixels
- NL_f = Incorrectly classified non-liver pixels

Both Segmentation Accuracy, and Liver Identification Accuracy were calculated for each of the 10 images. It is important to note that only the cone shaped B-Mode portion of the image was considered in segmentation accuracy, as otherwise the correct classification of surrounding black pixels would skew the results. Liver Identification Accuracy was considered as a second measure, since correctly classifying the liver is of greater interest for purposes of this project. However, it is trivial to achieve 100% Liver Identification Accuracy by simply classifying the entire image as liver. Thus, a high accuracy should be achieved in both measures to ensure that the algorithm is performing well. Initial segmentation tests were performed on one image at a time. If results were promising, the algorithm was applied to all 10 images and an average of each measure was calculated to determine overall performance of the algorithm.

C. Window Analysis

Simple intensity thresholding techniques can be easily performed on the entire image, as the relationship between neighbouring pixels is not considered. However, to determine effective thresholds for Noise, Entropy, Standard Deviation and Horizontal Autocorrelation (*listed in Section II.D. Segmentation Algorithms*), one must first segment the image into windows. The liver is a large and relatively homogeneous organ, consuming approximately 30% - 60% of the ultrasound image. Thus, it is assumed that texture features within liver windows are closely related, and distinct from non-liver windows. Quantitative texture values from Noise, Entropy, Standard Deviation and Horizontal Autocorrelation can be calculated for each window, and used as a means to segment the image.

Images were divided into windows, of a user-selected size. An example of window division can be seen in Figure 3 below. Note that only windows overlaying the B-Mode Image were considered.

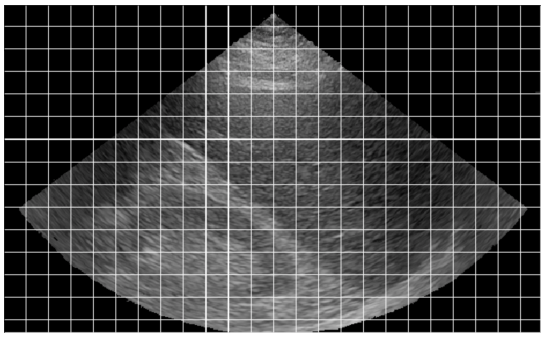

Fig. 3. Example of window division in ultrasound image. Window Size: 25x25 pixels.

For analysis purposes, windows in the ground truth liver and non-liver portions of the image were labelled. An example of liver and non-liver windows is presented in Figure 4 below. All Noise, Entropy, Standard Deviation and Horizontal Autocorrelation features were calculated for liver and non-liver windows and compared in histogram formats. Further details are presented in section *III. Results and Discussion*

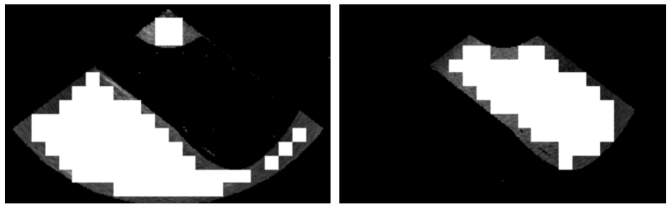

Fig. 4. Example of window division for non-liver (left) and liver (right) portions of the image, used for histogram analysis and threshold determination. Window Size: 25 x 25 pixels.

D. Segmentation Algorithms

1) Canny Edge Detection: Edge detection is a technique that utilizes changes in the brightness function of an image and marks them. With different order of derivatives, there exists several types of edge detection algorithms that each offer their pros and cons in the context they are applied for. The focus of this investigation will be on second order derivatives, as they measure the rate which the slope of the intensity surface changes (ie. concavity) with displacement. Second order derivatives are useful because they allow an easy interpretation of edge points, as they will contain a zero

crossing at the second derivative, allowing relatively easier localization of zero crossings [6]. Second order derivatives also change the sign of the computed magnitude at the centre of the edge.

The Canny edge detection algorithm, developed by John F. Canny in 1968, is composed of several layers of operations, allowing it to capture finer and wider ranges of edges in images. Another set of criteria that motivated the use of Canny edge detection for fatty liver segmentation were:

- Good detection; The Canny edge detection algorithm is able to locate and mark all real edges [7].
- Good localisation: The algorithm's output detected edge has minimal distance change from the true edge [7].

The multiple phases of Canny edge detection include:

i. Noise Reduction - Noise reduction is achieved by a radially-symmetric 2D Gaussian filter. The Gaussian filter is represented as follows:

$$G\theta(x, y) = e^{(r2/22)}$$

Where

$$r^2 = x^2 + y^2$$

The Gaussian filter utilizes a 5x5 sized kernel, with a standard deviation (sigma) of 0.4. The selection of the standard deviation value was based on the noise-smoothing trade-off, as with the increase of , noise will be suppressed the weakest edges will be smoothed away. The value of 0.4 was selected on the basis that the ultrasound images contain a wide range of fine detailed pixel segments.

ii. Intensity Gradient Computation - The purpose of computing the intensity gradient is to detect the edge gradient and direction. The purpose of doing so is the nature of edges existing in blurred images, which can be in horizontal, vertical, or diagonal orientation.The angles to be considered are 0, 45, 90, and 135. The gradient will be computed using the Sobel operator, and the angle will be computed through the tan inverse of the X and Y Sobel magnitudes. The X and Y Sobel matrices are as follows:

$$Sobelx = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix} \; Sobely = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

iii. Non-Maximum Suppression - Non-maximum suppression obtains points where the gradient magnitude is at a maximum along the direction of the gradient. The purpose of non-maximum suppression is to lessen the thickness of an edge to preserve edge uniformity throughout.

iv. Hysteresis Thresholding - Due to the inability to specify a threshold where the transition of an intensity gradient corresponding to an edge and non-edge, hysteresis thresholding is utilized. Hysteresis presents a spectrum of thresholds to detect edges, due to it setting a low and high thresholds, and with these thresholds. The in-between area considers its neighbours iteratively then declares it as an edge pixel it is connected to an edge pixel directly or via pixels lying above the maximum threshold. Figure 5 shows the regions of a sample image that contains variables that are above the maximum threshold value and ones that are in the in-between region.

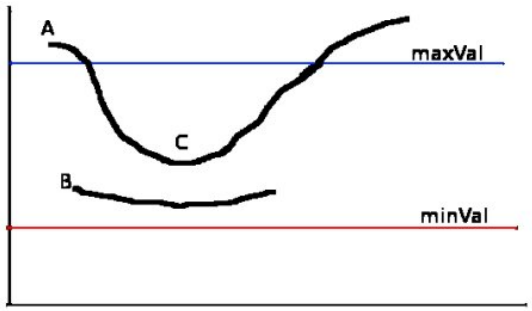

Fig. 5. Hysteresis Thresholding using Maximum, Minimum, and In-between Regions [8]

The portion of line "A" lying above the maximum threshold is an edge, and any other line below minimum is considered a non-edge. For the in-between region, any lines connected to the previously defined edge section is considered to be an edge. Any other lines in the in-between region that are not connected to the confirmed line ("A" in this example), are discarded.

The assumption of edges of interest being along a continuous curve allows following sections of the faint line, and discarding noisy pixels that do not constitute to a line but have produced large gradients.

2) Intensity Thresholding: Intensity thresholding is a simple image segmentation method, in which a histogram is first visualized and intensity thresholds between 0 and 255 are selected [6]. The selected thresholds correspond to troughs in the image histogram, which is ideally bi-modal or multi-modal. Histograms of the liver images will be analyzed, and fixed thresholds will be selected. Blurring and contrast enhancement techniques are applied prior to intensity thresholding. Darker intensities will be selected to segment the liver, thus including all pixels with intensities $<$ T (threshold). A mask for intensity thresholding is produces according the following equation [6]:

$$m(x, y) = \begin{cases} 0, & i(x, y) \geq T \\ 1, & i(x, y) \leq T \end{cases}$$

Where:

- $i(x, y)$ = Original Image
- $m(x, y)$ = Liver Segmented Mask
- T = Threshold

In addition, Niblack's thresholding was explored. Niblack's thresholding is an adaptive thresholding method in which considers the mean and standard deviation within windows of the image. The threshold T is dynamic according to the following equation [9].

$$T = \mu + K * S$$

Where:

- T = Threshold
- S = Standard Deviation
- μ = Mean
- K = A constant dependant on noise

Histograms only take into account the intensity values of individual pixels, however they do not take into account the spatial arrangement of pixels. Valuable information can be extracted from the texture of an image. Thus, entropy thresholding, noise, standard deviation, and horizontal autocorrelation are considered in the following sections.

3) Noise Detection: Using windows described in section *II.C Window Analysis* fast noise estimation was applied. This method estimates the variance of additive Gaussian noise. A 3x3 filter is used and calculates the standard deviation of the noise. The 3x3 filter is displayed below:

$$\text{Filter} = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

$$\sigma = \frac{\sqrt{0.5 * \pi}}{6 * (W-2) * (H-2)} * \sum\sum Filter * ImageWindow$$

Where:

- W = Window Width
- H = Window Height

A value of σ greater than 10 typically indicates that the image is very noisy [6]. Fast noise variance is a no-reference metric, and thus its scores may not always correlate with what is observed subjectively. However it is applicable as segmentation is based on window to window comparisons and threshold selection.

4) Entropy Thresholding: Using windows described in section *II.C Window Analysis* entropy calculations were applied. Shannon entropy was calculated to each window according to the following equation:

$$S = -\sum(pk * log(pk))$$

Where:

- S = Shannon Entropy
- pk = frequency/probability of pixels of value k

Shannon Entropy is useful for both edge detection and texture classification [10]. Results of windowed Shannon Entropy are used to determine thresholds for liver and non-liver segmentation.

5) Standard Deviation Thresholding: Using windows described in section *II.C Window Analysis* standard deviation calculations were applied. An estimate of standard deviation of pixel intensities within a window, provides a quantitative measure for the amount of variation. It is yet another texture analysis technique which may be useful for liver / non-liver window identification. A simple cv2.meanStdDev function was applied to each window to calculate standard deviation. Based on histogram analysis, and appropriate standard deviation threshold can be selected.

6) Horizontal Autocorrelation: Autocorrelation is a function of distance and direction of separation, which compares pixel pairings and the likelihood that they will have the same intensity[11]. Horizontal autocorrelation was selected as the tissue fibres are aligned in a horizontal direction, and appear to be somewhat different between liver and non-liver tissues. With an offset of 4 pixels, the horizontal autocorrelation function used was as follows:

$$S = \sqrt{\sum window(x,y) * window(x,y+4)}$$

Essentially, the autocorrelation is result of convolving a function with itself. Based on histogram analysis, and appropriate horizontal autocorrelation threshold can be selected to differentiate between liver and non-liver windows.

E. Process

To successfully segment the liver, it was planned to use a combination of edge detection, intensity thresholding, and texture thresholding methods. Ideally, the edge detection algorithms will create borders around distinct objects within the image. For example, the kidney, seen as an oval shape in the bottom left of Figure 1, the diagonal line defining the edge of the liver, and the bright layers of skin found at the top of the image. Evidently, an edge detection algorithm will not classify which object is which. Thus, the borders will be used to define shapes, in which intensity and texture thresholding techniques can be applied. If the majority area, or majority windows in an object meet a certain threshold (ie. entropy > 6), the object can be classified as liver, kidney or epidermal tissue. This requires thorough experimentation and image analysis, to determine optimal thresholds, and combinations of thresholds across various texture features that accurately define the tissue.

III. Results and Discussion

1) Canny Edge Detection: The Canny edge detection algorithm was tested two times on the healthy and fatty liver images. The first test was through a manual construction of the algorithm, and the second test was using the algorithm built in the OpenCV library. The purpose of using the Canny edge detection algorithm is to interpret the differences between the sparsity of healthy and fatty liver. Figure 6 shows the result of implementing the Canny edge detection algorithm to a healthy liver image. The anticipated outcome was a distinctive set of edges, and the goal of that was to obtain the separation (shown in the image on the left) between the liver and kidney regions.

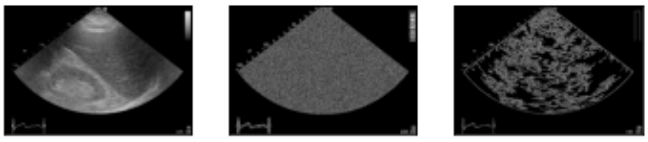

Fig. 6. Healthy liver (left); manually constructed Canny edge detector (middle); Canny edge detector using OpenCV library (right)

The expected result was lesser edges for the liver and more distinct edges to be output in the kidney region, but that

was not the case. The outcome, however, was fine detection of all the edges on the image, which also included some noise. This outcome is not useful to segment the liver from the ultrasound scans, as no useful information regarding the diagonal distinguishing line have been identified. After further investigation, the failure to detect the edges of liver, kidney, and related regions can be attributed to the fact that the algorithm follows a faint section of a given line, discarding noisy pixels that do not constitute to an edge line, but have produced large gradients The outcome anticipated initially was to be similar to what Cao et. al. had when Canny edge detection was applied to liver cirrhosis ultrasound images [12]. The result Cao et. al. had was sparse edges for liver cirrhosis compared to healthy liver. After further examination of the ultrasound scans, fatty liver seemed to more blurred than a healthy liver ultrasound scan,so a distinct difference between the two images, in terms of edge features, can not be extracted compared to a liver cirrhosis case, where the unhealthy liver is very distinctly different than a healthy live image. Figure 7 shows the result of applying the Canny operator on both healthy and fatty livers, and it can be seen the the difference in the region of interest, the liver, kidney, and the separating diagonal tissue are not detected as anticipated.

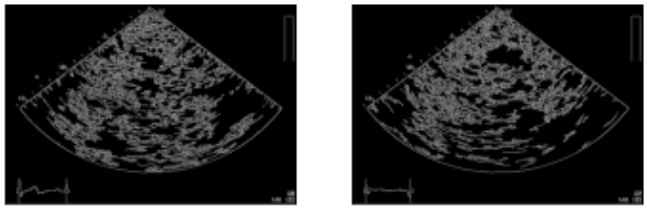

Fig. 7. Edges of healthy liver (left) versus fatty liver(right)

To accentuate the edge features in the pre-Canny operated images, contrast adjustment was performed via histogram equalization. Figure 8 shows the result of applying histogram equalization to the ultrasound image prior to inputting the image to Canny operator, and it can be seen that histogram equalization worsens the quality of the output image for the required context. Most of the detail from the pre-operated image, including potentially noise, have been detected.

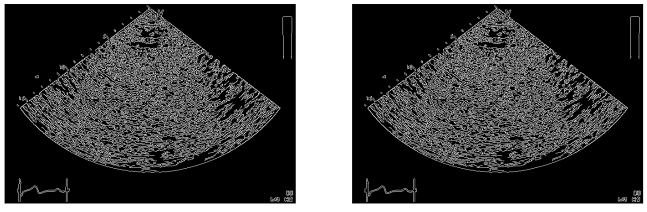

Fig. 8. Edges of healthy liver (left) versus fatty liver(right) implementing Histogram Equalization

The conclusion made after implementing the Canny operator is that this operator would be excellent in detecting smaller and finer details, in images containing minor details to be extracted, satellite images for example. However, the Canny operator would not be an appropriate choice to select the regions of interest which are the liver, kidney, and diagonal separating tissue. A more appropriate edge detection algorithm would be one that detects global objects, compared to more local edge detection that was believed to be more effective prior to conducting the experiment.

2) Intensity Thresholding: Simple Intensity Thresholding: Initially, simple thresholding techniques were applied based on histogram analysis. Histograms were plotted only for intensities 1-255. Intesnity value 0 was disregarded as the black space in the image does not represent the B-Mode ultrasound. Plotting the histogram of Image 1, no distinct troughs were found as seen in Figure 9.

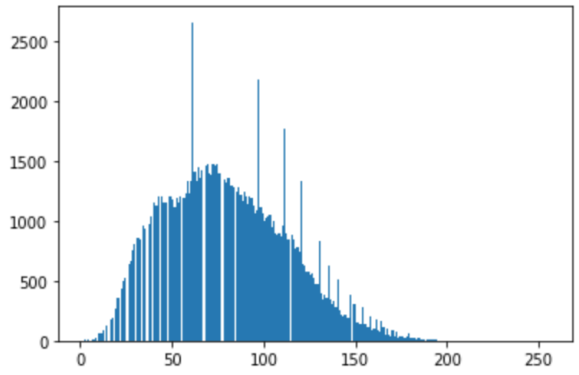

Fig. 9. Histogram of Image 1 (no filtering applied)

Ultrasound images are prone to noise, which affect the shape of their histograms. Therefore, several filtering techniques were applied including:

- Mean Filter Size 5x5
- Mean Filter Size 11x11
- Gaussian Filter Size 5x5
- Centre Weighted Median Filter

Histograms were re-plotted for the images, in search for the filter which would result in the most significant troughs. It was determined that the Mean 5x5 Filter and Gaussian filters produced the best histograms. Figure 10 below shows the resultant histogram after a mean 5x5 filter was applied to Image 1. For detailed analysis and explanations of trial and error methods, please refer to the submitted Jupyter Notebook.

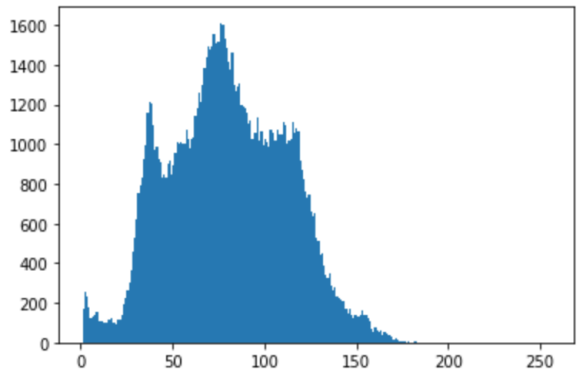

Fig. 10. Histogram of Image 1 after 5x5 means filtering was applied

Although the troughs are not significant, they provide guidance for thresholds that may help in the liver segmentation process. It is inevitable that there will be some misclassification at this stage. Initially, thresholds of 25, 50 and 90 were applied. Through qualitative analysis and comparison to the ground truth mask, as well as quantitative accuracy results, it was determined that a fixed threshold of 85 was best for Image 1. The result of thresholding is displayed in Figure 11 below.

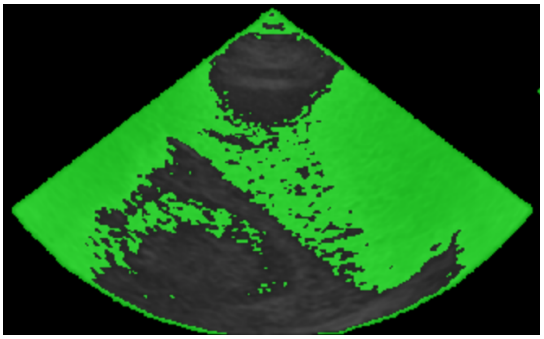

Fig. 11. Threshold Segmentation of Image 1 with Thresholding ¡85

Most of the liver is captured, however some areas which are clearly non-liver (such as the kidney and dark shadowing effects on the left side of the image) are misclassified. For Image 1, the quantitative results were:

- Segmentation Accuracy: 72.82%
- Liver Identification Accuracy: 85.58%

Therefore, fixed thresholding is a promising segmentation technique. When applied to all 10 images, the threshold 85 did not function well. Averaging between the 10 images, with a threshold of 85, the quantitative results are:

- Segmentation Accuracy: 55.63%
- Liver Identification Accuracy: 77.01%

Evidently, a single threshold does not hold true for all images. Its also worth noting that the thresholding technique worked better on the 3 images classified as healthy. The healthy liver is darker than the kidney and is segmented well as seen in Figure 12 below.

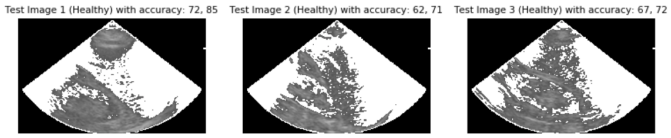

Fig. 12. Threshold Segmentation of Healthy Images 1-3 with Thresholding ¡85

The remaining 7 images are from patients diagnosed with fatty liver disease, and thus their livers appear brighter. They are as bright, or brighter than borders of the kidney and portions of skin, making it much harder to segment. See Figure 13 below.

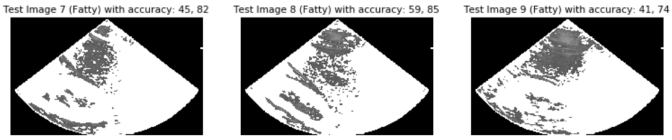

Fig. 13. Threshold Segmentation of Diseased Images 7-9 with Thresholding ¡85

Through trial and error, it seemed that the only way to combat the fatty liver misclassfication was to decrease the threshold for diseased livers and capture only the darkest portions. In this way, a lot of the liver is missed, but less non-liver is falsely classified as liver (improving overall segmentation accuracy). The results after decreasing the threshold for fatty livers to 75 were:

- Segmentation Accuracy: 57.25%
- Liver Identification Accuracy: 67.90%

Contrast enhancement through histogram equalization was also applied, with the intention of creating a more significant intensity difference between the liver and non-liver. Unfortunately, contrast enhancement produced significantly worse results. The shadowing artifacts in the ultrasound images were exaggerated and the resultant liver tissue contained two very different intensities. Figure 14 below displays an example of an image where contrast enhancement produced worse segmentation results.

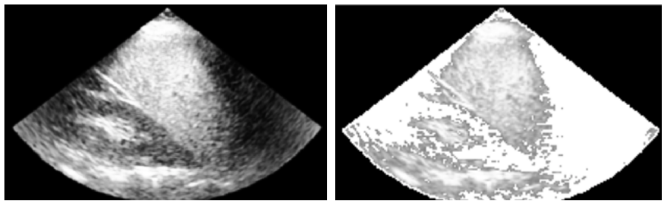

Fig. 14. Sample Image 4 presenting the negative effects of contrast enhancement for Intensity Thresholding

Thus, contrast enhancement through histogram equalization was abandoned. With upper and lower bounds of 10-85 and 10-75 for healthy and fatty images respectively the best results were:

- Segmentation Accuracy: 57.25%

- Liver Identification Accuracy: 67.90%

Niblack's Intensity Thresholding: Niblack's Intensity Thresholding, as predicted, was not useful for liver segmentation. As the algorithm uses local thresholds rather than a single global threshold, it exaggerates texture features, rather than segmenting large object like the liver. Various window sizes were selected however the results were poor, and thus no further analysis was done in this section. For reference, Figure 15 below presents the failure of the algorithm for this application.

Fig. 15. Sample Image 1 presenting the non-ideal result of Niblack's Thresholding

3) Noise Detection: After segmenting the image into windows, as per *II.C. Window Analysis*, fast noise estimation was applied to each window. The noise for ground truth liver and non-liver segments were displayed on a histogram. As seen in Figure 16, there is no significant difference in the distributions for liver (blue) and non-liver (orange) windows.

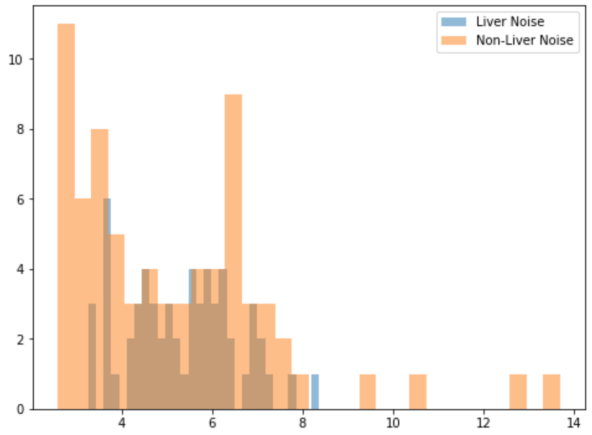

Fig. 16. Distribution of Fast Noise Estimation in Liver and Non-Liver Windows for Image 1

Furthermore, there was no significant difference in the mean noise:

- Mean Noise in Liver: 5.33
- Mean Noise in Non-Liver: 5.14

This was tested on all 10 images, and window sized of 25 and 11. No consistent differences in distributions were noted. Therefore, segmentation through noise was abandoned as a technique.

4) Entropy Thresholding: Entropy Thresholding produced promising results. Results of Shannon entropy in size 25x25 windows of liver and non-liver tissue are displayed in Figure 17.

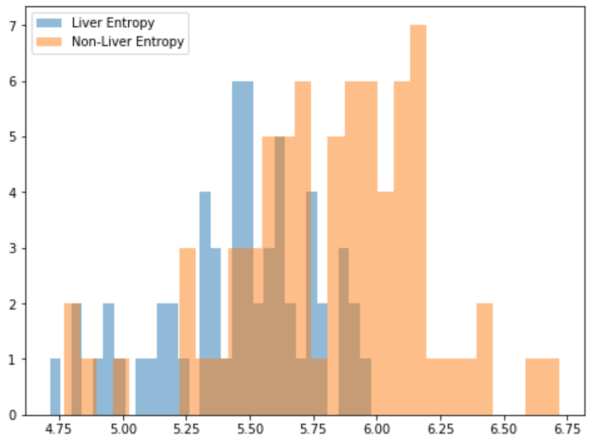

Fig. 17. Distribution of Shannon Entropy in Liver and Non-Liver Windows for Image 1

From the histogram, one can clearly see that any windows with a Shannon Entropy $>$ 6.00 are certainly non-liver. These non-liver windows in Image 1 are presented in Figure 18 below.

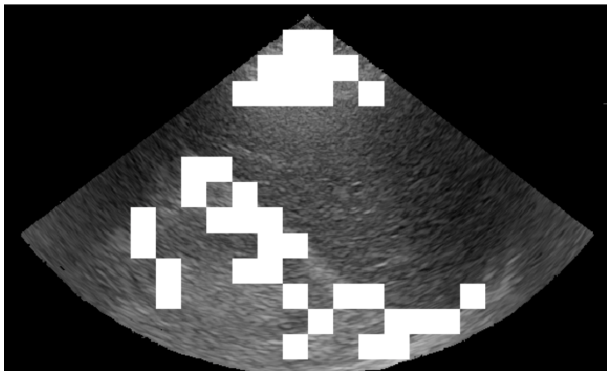

Fig. 18. Windows with Shannon Entropy ¿ 6.00 - successful technique to identify non-liver tissue

Entropy thresholding was then tested on each of the 10 images individually. In all 10 images, a threshold between 5.8 and 6.1 accurately defined several windows which are certainly non-liver.

5) Combination of Intensity Thresholding and Entropy: Since entropy detection proved to be useful in detecting non-liver areas, and intensity based thresholding misclassified several non-liver areas as liver, it was predicted that a combination between the two algorithms would result in improved liver segmentation. An algorithm was written which takes the mask resulting from intensity thresholding, and sets any non-liver areas identified by the entropy algorithm to 0. In addition, it was observed that many of the windows detected by the entropy algorithm as non-liver were in the lower left portion

of the image, where kidney was often classified (especially in fatty liver images as per section *III.2. Intensity Thresholding*). Using prior knowledge - that the location of the kidney is always in the bottom left of the image - it would be useful to eliminate all pixels below, and to the left of high entropy windows (excluding those at the top of the image which represent skin). This may also provide an opportunity to increase the intensity threshold for fatty liver images again, and capture more liver tissue. To summarize, the combination algorithm written does the following:

- a. Takes the resultant mask from intensity thresholding as an input.
- b. Calculates the Shannon Entropy value for all 25x25 windows on the original B-Mode image.
- c. Classifies all windows with a Shannon Entropy > 5.9 as non-liver.
- d. If any pixels were classified as liver by intensity thresholding, and non-liver by entropy detection, they are re-classified as non-liver.
- e. For any non-liver entropy windows below pixel height 150 (which is certainly below the skin), all pixels below and to the left of the window are re-classified as non-liver.

The results of this algorithm, with an intensity threshold of 85 and entropy threshold of 5.9 were impressive. Across all 10 images, the average results were as follows:

- Segmentation Accuracy: 73.13%
- Liver Identification Accuracy: 74.85%

This is a >15% improvement in average segmentation accuracy and >6% improvement in Liver Identification. The images with the worst and best results are displayed in Figure 19 below.

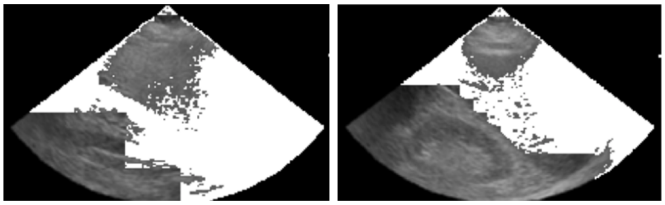

Fig. 19. Worst (left) and Best (right) images resulting from the Thresholding & Entropy Combination Algorithm

The images had segmentation accuracies of 58% and 85% for worst and best respectively. 7/10 images were achieved a segmentation accuracy >71%. In observing the images qualitatively, it appears the liver identification accuracy could improve by implementing a closing function, as there are several small holes in areas classified (correctly) as liver. After implementing a closing morphology function with window size 11x11 to all resultant masks, accuracies improved as follows:

- Segmentation Accuracy: 75.17%
- Liver Identification Accuracy: 84.05%

This is a >2% increase in segmentation accuracy and >9% increase in liver identification accuracy. Updated results from the worst and best images are displayed in Figure 20 below.

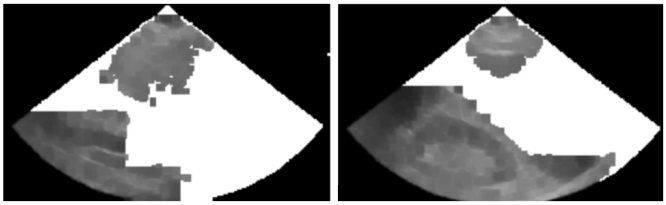

Fig. 20. Worst (left) and Best (right) images resulting from the Thresholding & Entropy Combination Algorithm after Morphology Closing was applied

6) Combination of Intensity Thresholding and Standard Deviation Thresholding: In a manner very similar to entropy thresholding, the standard deviation in 25x25 windows showed promising results for identifying non-liver windows. Following an identical process to section *III.4. Entropy Thresholding* a universal standard deviation threshold for classifying non-liver images was determined to be 19.5. An algorithm, nearly identical to the intensity threshold and entropy combination algorithm was written, replacing entropy with standard deviation detection. The algorithm has the followeing steps:

- Takes the resultant mask from intensity thresholding as an input.
- Calculates the standard deviation value for all 25x25 windows on the original B-Mode image.
- Classifies all windows with a Standard Deviation > 19.5 as non-liver.
- If any pixels were classified as liver by intensity thresholding, and non-liver by standard deviation detection, they are re-classified as non-liver.
- For any non-liver standard deviation windows below pixel height 150 (which is certainly below the skin), all pixels below and to the left of the window are re-classified as non-liver.

After applying closing morphology to the results of the intensity threshold and standard deviation thresholding algorithm, the average results are as follows:

- Segmentation Accuracy: 79.43%
- Liver Identification Accuracy: 86.83%

This is yet another improvement. From the intensity entropy combination algorithm, standard deviation results in >4% improvement in segmentation accuracy and >2% improvement in liver identification accuracy. The non-liver windows defined by standard deviation are thus more representative than those defined by entropy. The results listed above were the highest achieved for purposes of this assignment. All 10 images achieved a segmentation accuracy >71%. Results from the worst and best images are displayed in Figure 21 below.

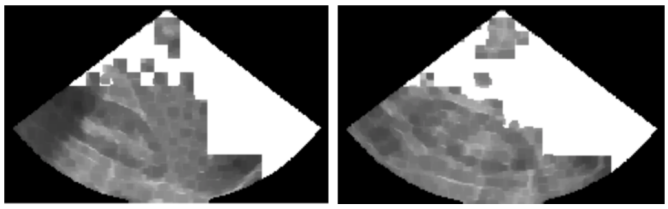

Fig. 21. Worst (left) and Best (right) images resulting from the Thresholding & Standard Deviation Combination Algorithm after Morphology Closing was applied

The images had segmentation accuracies of 71% and 87% for worst and best respectively. It is clear that the worst image failed as a result of windows in the middle of the liver, and higher than pixel 150 which were classified as non-liver by the standard deviation algorithm. This is the only image of the 10 which suffered from this problem. As a result, all pixels below and to the left of the non-liver windows were also wrongly re-classified as non-liver. It is inevitable that the threshold for standard deviation does not hold perfectly for all 10 images.

7) Horizontal Autocorrelation: Horizontal Autocorrelation Thresholding produced inconsistent results. Results of calculating horizontal autocorrelation with and offset of 4 pixels in size 25x25 windows of liver and non-liver tissue are displayed in Figure 22.

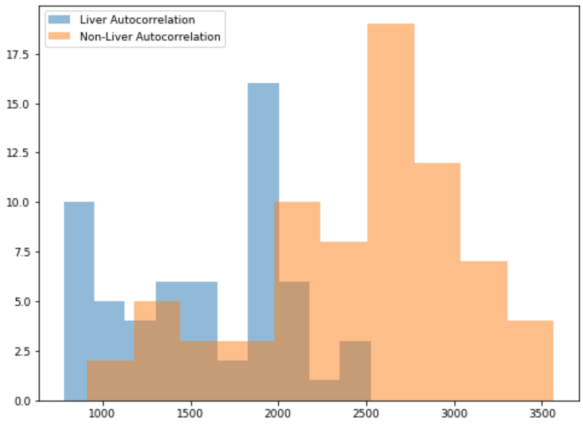

Fig. 22. Distribution of Horizontal Autocorrelation Values in Liver and Non-Liver Windows for Image 1

From the histogram, one can clearly see that any windows with a horizontal autocorrelation value $>$ 2500 are certainly non-liver. These non-liver windows in Image 1 are presented in Figure 23 below.

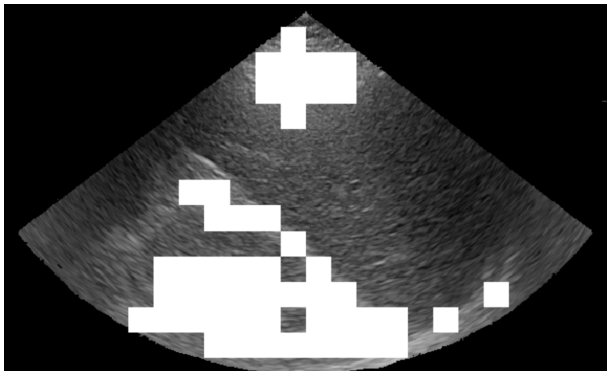

Fig. 23. Windows with Horizontal Autocorrelation ¿ 2500 - successful technique to identify non-liver tissue in Image 1

Horizontal autocorrelation was then tested on each of the 10 images individually. Unfortunately, thresholding results were extremely inconsistent. In some images, the histograms were almost entirely overlapping. See Figure 24 below for an example.

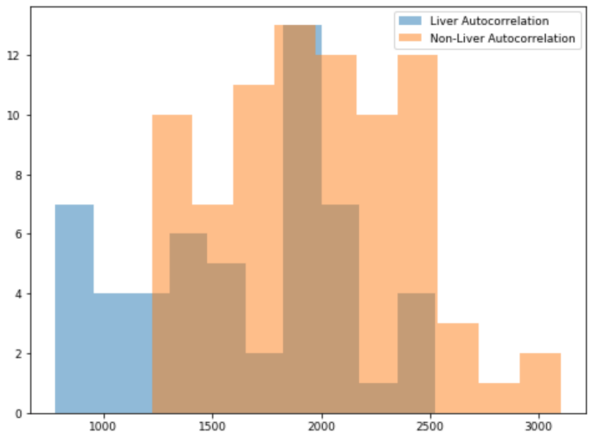

Fig. 24. Distribution of Horizontal Autocorrelation Values in Liver and Non-Liver Windows for Image 6 showing inconsistency

Due to these inconsistencies, no universal threshold was selected and horizontal autocorrelation was not further pursued for liver segmentation. It is possible that with further testing, varying window sizes and varying pixel offsets that the horizontal autocorrelation feature may be useful, however due to time constraints it was not thoroughly explored.

IV. Conclusion

Ultrasound images for fatty liver detection are widely acknowledged as an efficient, accurate, and non-invasive method. The several image segmentation techniques investigated in this experiment resulted with varying success. Canny image detection algorithm resulted with unusable edges detected, due to the focus on local edges, opposed to a global edge detection algorithm which would be able to distinguish the liver, kidney, and the diagonal separation tissue.

To conclude, the best average results in liver detection were 79.43% and 86.83% for segmentation accuracy and

liver identification accuracy respectively. These results were achieved using a combination of intensity thresholding, windowed standard deviation calculations, and morphology closing techniques. Results can likely be improved if combined with successful edge detection techniques focused on global detection rather than localized details. There are also several further techniques for texture classification which can be thoroughly explored including wavelet based features, and features extracted from gray-level co-occurrence matrices.

REFERENCES

[1] M. Clinic. Nonalcoholic fatty liver disease. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/nonalcoholic-fatty-liver-disease/symptoms-causes/syc-20354567

[2] M. N. M. F. Ajit Ramakant Mahale, Sonali Dattatray Prabhu and S. Ullal, "Clinical relevance of reporting fatty liver on ultrasound in asymptomatic patients during routine health checkups," *J Int Med Res*, vol. 46, no. 11, p. 4447–4454, 2018.

[3] A. Jepson. Image segmentation. [Online]. Available: http://www.cs.toronto.edu/ jepson/csc2503/segmentation.pdf

[4] J. A. Noble, "Ultrasound image segmentation and tissue characterization," *Proc Inst Mech Eng H.*, vol. 224, no. 2, pp. 16–307, 2010.

[5] "Data-set of b-mode fatty liver images." [Online]. Available: https://www.kaggle.com/shanecandoit/dataset-of-bmode-fatty-liver-ultrasound-images

[6] D. D. Nikitenko, "Lecture notes fall 2020." [Online]. Available: https://moodle.socs.uoguelph.ca/course/view.php?id=173

[7] U. Bagci. Canny edge detection. [Online]. Available: http://www.cs.ucf.edu/ bagci/teaching/robotvision18/Lec6.pdf

[8] "Canny edge detection." [Online]. Available: https://docs.opencv.org/master/da/d22/tutorial$_p y_c anny.html$

[9] S. N. Efficient implementation of niblack thresholding for mri brain image segmentation. [Online]. Available: http://www.ijcsit.com/docs/Volume%205/vol5issue02/ijcsit20140502274.pdf

[10] B. Singh. Edge detection in gray level images based on the shannon entropy. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?d10.1.1.165.9458

[11] C. Robertson. Theory and practical recommendations for autocorrelation-based image correlation spectroscopy. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3414238/

[12] B. H. Guitao Cao, Pengfei Shi1, "Liver fibrosis identification based on ultrasound images," *Engineering in Medicine and Biology*, 2005.