Abstract—The presented paper demonstrates how to detect pathological issues in transcranial B-MODE ultrasound images. We developed an algorithm based on binary thresholding with subsequential computing of area inside ROI which represents an area in which we detect the issue. We work with a collection of images acquired from 3 different ultrasound machines. We detect echogenic area in substantia nigra which is primary feature to Morbus Parkinson and also raphe nucleus echogenicity to detection of other neurological diseases. All achieved results were verified by an erudite neurosonologist and based on statistics such as correlation and kappa analysis. Average correlation between observers $r > 0.88$, level of agreement kappa $> 0.82$. So, we proved that developed algorithm is highly reproducible and also could be used for different cases, not only in neurology due to principle of B-MODE imaging. The algorithm has been implemented in MATLAB with Image Processing Toolbox. Achieved results will verified for 3 different ultrasound machines.

Keywords—B-MODE, Parkinson, echogenicity, hyperechogenic SN, substantia nigra parkinson

I. INTRODUCTION

Imaging methods and their computerized analysis is one of the inseparable parts of modern medicine. Diagnostic ultrasound is a non-invasive imaging method on which is focused this research. Diagnostic ultrasound is safe, non-invasive and repeatable method with no negative radiation. This paper is focused on processing of B-MODE transcranial images. Each structure in B-MODE image is characterized by its echogenicity level which represents tissue density more respectively. Principle of B-MODE is based on different echogenicity of tissues which means tissues with different density are displayed with different intensity which represent tissue echogenicity. So that, this displaying is universal for different structures because the principle is the same for displaying brain structures, cardiovascular images, etc. Details about B-MODE and ultrasound physics are available in [4].

A. B-MODE background and using in neurology

In neurology is B-MODE used for different diagnoses and is highly reproducible. This paper is focused on using B-MODE to detection of Morbus Parkinson feature and features for other neurological diseases such as panic disorders, multiple system atrophy, etc. B-MODE images are natively in grayscale. In other words, in B-MODE is each structure (tissue) characterized by brightness level $I \in (0, 255)$ which represents echogenicity, tissue density more respectively. Principle of B-MODE is based on different echogenicity of tissues which means tissues with different density are displayed with different intensity which represent tissue echogenicity. So that, this displaying is universal for different structures because the principle is the same for displaying brain structures, cardiovascular images, etc. Details about B-MODE and ultrasound physics are available in [4].

B. Examined substantia nigra and raphe nuclei in B-MODE

B-MODE is able to perfect visualize echogenic signals, hyperechogenicity more precisely. In case of Morbus Parkinson (Parkinson’s Diseases; PD) we analyze hyperechogenicity of ipsilateral and contralateral substantia nigra (SN) structure [2], [3], [9]–[13], [16] as the parkinsonism feature in B-MODE. In case of raphe nuclei (RN) is hyperechogenicity considered as normal feature. So that, in case of substantia nigra is pathological hyperechogenicity (we do not distinguish 5 grades of echogenicity which are established) and in case of raphe nuclei is pathological black area (no echogenic structure found) [13]. The following figure shows the position of both of the investigated structures.

This work was supported by the operational program “Centrum excellence IT4Innovations” number CZ.1.05/1.1.00/02.0070 supervised by Assoc. Prof. Petr Sosík for Silesian University in Opava.

2 http://www.criticalecho.com/content/tutorial-2-modes-ultrasound
4 http://www.bioportfolio.com/resources/jarticle/79341/Dementia-In-Morbus-Parkinson-Reasonable-Diagnostics-And-Rational-Therapy.html
5 http://www.neuroanatomy.wisc.edu/virtualbrain/BrainStem/20Substantia.html
II. THE DEVELOPED ALGORITHM

We developed the algorithm focused on ROI-based area measurement of echogenicity. It is based on computing of the area inside ROI. We selected MATLAB with Image Processing Toolbox\(^8\) and MATLAB Compiler\(^9\) to deploy stand alone applications. MATLAB with this toolbox provides a wide range of tools and functions which we need and we created fully GUI-based application. Main reasons why MATLAB has been selected:

- Working with medical DICOM images
- A lot of built-in functions to image analysis – (histogram, ROI, layers, etc.)
- Creating GUI-based stand alone application for practical using in practice
- Visual design of GUI

It is not necessary to solve “routines” such as how to plot the histogram, how we select ROI as a layer, etc. but solve the core of this research. Also support of DICOM which is a worldwide standard\(^10\) of imaging, storing and sharing of medical image data, is very important but the majority of used images is in DICOM format.

In this work we use images acquired from 3 ultrasound machines – Philips HDI 5000 (JPEG output only), ESAOTE My Lab and GE Vivid 7 (both of them with raw DICOM output). All achieved results are compared and evaluated for all mentioned machines. Simultaneously we have data from other device but it is not completed up to the present.

A. Brief description

The developed algorithm is primarily based on binary thresholding. By means of the binary thresholding we simply compute pixels which are inside ROI after thresholding. We consider \(H \in [0, 255]\) as the intensity level of a pixel and \(T\) as the threshold; the idea of the binary thresholding is simple:

\[
\text{IF } T \leq H \text{ THEN } \text{pixel_black ELSE pixel_white (1)}
\]

This algorithm computes thresholding for all \(T \in [0, 255]\) and also computes number of white pixels. The number of white pixels decreases with increasing \(T\). This number represents the area inside ROI which is automatically converted into real \(\text{mm}^2\) depending on image resolution according to used ultrasound machine.

B. Parts of the developed algorithm and pre-processing

The algorithm is composed from parts which are described in our previously published papers [9], [10], [11], [12] in details.

The parts of the algorithm are the following:

\(^7\) http://www.ncbi.nlm.nih.gov/pubmed/19219216
\(^8\) http://www.mathworks.com/products/image/
\(^10\) http://medical.nema.org/
• Cut window of 50×50 mm, automatically converted into grayscale if it is necessary
• Automatic global average intensity detection against false diagnosis
• Select needed ROI
• Run the area computing
• Representation in the curve with option of saving results

From input image (Fig. 1) we cut the window 50×50 mm which is needed to the processing. This process is automatic or manual by click on 2 points on the native axis of ultrasound image, see Fig. 2. From the points with coordinates \((x_1, y_1)\) and \((x_2, y_2)\) is computed Euclidean distance

\[
d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}
\]

and we get 50×50 mm window. The native axis represents real dimensions in mm (0, 5, 10,…) as follows.

![Fig. 4 Window of 50×50 mm which we need to the main processing](image)

Despite B-MODE images are natively in grayscale, can be stored as 24bit RGB; for example the image can contains color items such as head, color axis, etc. Selected window of 50×50 mm is always gray but for computer is stored as RGB. So that, each pixel is represented by RGB vector \(p_{RGB} = (R, G, B)\) in which \(R\), \(G\) and \(B\) represent values of RGB channels. Let \(H \in \{0, 255\}\) we need to find a function which realizes \(p_{RGB} \rightarrow p_H\) for each image pixel. In grayscale is \(R = G = B \Rightarrow H\).

\[
H = 0.299R + 0.587G + 0.111B
\]

which realizes \(p_{RGB} \rightarrow p_H\) for each image pixel. In grayscale is \(R = G = B \Rightarrow H\).

C. Quality image control as a multi-agent based part

It is a very useful part of the algorithm against false echogenicity detection and it is independent on the main processing. The part is realized as a simple multi-agent system [11].

If the input image has high minimal level of brightness could be incorrectly evaluated because very light areas could be recognized as echogenic but in fact are not significantly echogenic. The following figure shows the example of an unsuitable input image and a normal image.

![Fig. 5 Unsuitable input (left), normal image (right)](image)

We divided the window (Fig. 4) into blocks of 16×16. Each block represents the average intensity value of pixels which are inside block. [11], [12]. We get blocks \(B_1, B_2, \ldots, B_{256}\) and their intensities \(IB_1, IB_2, \ldots, IB_{256}\). The algorithm computes minimal intensity

\[
\min(IB_1, IB_2, \ldots, IB_{256})
\]

Fig. 6 Block 16×16 to check minimal intensity

We depended on experience of neurosonologist only but we need a proof that this value is really right value how to separate normal and corrupted images. In the previously presented version of the algorithm we set a priori value of 25. In other words, if the computed minimum (2) exceeded 25, the image has been considered as corrupted in consequence of high minimal intensity. Informally, if the minimal intensity (2) is too high, “black is not visually really black”. The value of 25 has been determined by neurosonologist.

In previously presented version of the algorithm we set a priori value of 25. In other words, if the computed minimum (2) exceeded 25, the image has been considered as corrupted in consequence of high minimal intensity. Informally, if the minimal intensity (2) is too high, “black is not visually really black”. The value of 25 has been determined by neurosonologist.

We depended on experience of neurosonologist only but we need a proof that this value is really right value how to separate normal and corrupted images. In the previously

published version we computed this value from histograms but the algorithm was not changed.

We work with images from different ultrasound machines and each of them could have different settings. This value has been determined from histogram analysis of the window 50×50 mm (Fig. 2). We consider the window and its histogram which is represents total number of pixels in the window as a function of their intensity, formally is given by

\[ IM_k = \sum_{i=0}^{K} p_i \]  

(4)

which is the sum of all intensities. Histogram is our tool how to determine the border value. From the window (Fig. 2) we computed minimal, maximal and average value of intensity for all images from Philips, GE and ESAOTE, totally more than 500 histograms created. The average value is used descriptors to determination as the border value to image quality analysis (Fig. 4).

From histograms we computed average intensity for each image and from these values has been subsequently calculated general average intensity for Philips, GE and ESAOTE as follows:

![Average intensity values from histogram – Philips](image1)

![Average intensity values from histogram – ESAOTE](image2)

Fig. 7 Measured average intensity from histogram of the window of 50×50 mm

From this analysis we obtained the following results:

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Images</th>
<th>Ave. Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philips</td>
<td>100</td>
<td>25.4</td>
</tr>
<tr>
<td>ESAOTE</td>
<td>420</td>
<td>28.3</td>
</tr>
<tr>
<td>GE</td>
<td>68</td>
<td>27.8</td>
</tr>
</tbody>
</table>

Table 1 Observed average intensity from histogram

This analysis proved that initial value of 25 determined by neurosonologist was close to computed values but we needed this proof.

These obtained values were stored into the algorithm. If we load the input image, the program automatically detects device from DICOM info and sets computed value. If we load a new image, the algorithm uses default value of 25 and we can change it. We have data from other device on which we try this evaluation but it is not completed.

We also discuss about block size to global average intensity detection. The neurosonologist determined as optimal block size of 5×5 mm according to the resolution and window size [11], [12], [13]. If we used more blocks such as 32×32, very often was evaluated as minimal intensity \( H = 0 \) as in isolated black pixel and this is inapplicable. In case of using big block such as 8×8 blocks only, evaluated values were too inexacty because the block represents large area. In accordance with these facts we have selected 16×16 blocks as optimal size as for the window of 50×50 mm.

![Fig. 8 Grid of 8×8 incorrectly represents global intensity](image3)

If we use only 64 blocks, average intensity in each block is represented inexacty because the block contains too many pixels. In other words, larger blocks are less exact but too small blocks are too exact and could be computed isolated very small area with \( H = 0 \). Blocks of 16×16 are optimal and the size of this block was also verified by neurosonologist according to resolution and window size.

The agent-based system is designed as an isolated part of the processing which helps to distinguish normal and corrupted images. So, the multi-agent system has 4 agents as follows:

\[ \text{MAS}_{\text{quality}} = \{a_{\text{Philips}}, a_{\text{ESAOTE}}, a_{\text{GE}}, a_{\text{unknown}}\} \]  

(5)

Each agent compares minimal detected intensity (3) with the computed value for the ultrasound according to Table 1. This
multi-agent system can be extended if we define more values for other ultrasound machines. If the algorithm detects unknown machine never used before, will be set the default value of 25 (a unknown). In other words, each agent has stored the appropriate computed value which is compared with detected minimal intensity. Furthermore, we can extend this system with other new agents for new areas, window more respectively. In future, if we will observe some new area we can simply add a new agent for a new window which with stored computed minimal value. In other words, each agent will be optimized according to observed area.

Currently we work only with area a2 with visualized SN and RN.

This agent-based evaluation could be incorrect with the function Gray Scale Map12 which allows set many preset types of grayscale map according to tissues to the best subjective resolution and brightness, e.g. instead of the range (0, 255) is confined only to (50, 180) for the best visualization. It should be disadvantage for objective US image processing and echogenicity detection. Each image can be set with different gray scale. To avoidance of the described problems we use sets of images with the same gain settings with the range (0, 255).

D. The main processing and evaluation of pathology

The main processing of the presented algorithms is based on binary thresholding inside ROI as we mentioned.

To compute of the area inside ROI we must select an appropriate ROI. We work with different ROI which have different initial area depending on structure. We use elliptical ROI with 50 mm² of the initial area for substantia nigra and 30 mm² for RN [10], [11], [12], [13].

\[ \sum_{T=0}^{255} a_T \]

Decreasing speed is the primary marker to diagnosis pathology and physiology, see [11], [12], [13]. That depends on structure properties. The following figure shows 4 cases – normal SN, echogenic SN, normal RN and anechogenic RN.

![Fig. 10 The used ROI for RN (left) and for SN (right)](image-url)

The number in the graph represents summation of all computed values of the area. Let \( a_T \) represents the area for \( T \in (0, 255) \), then we obtain

\[ \sum_{T=0}^{255} a_T \]

12 http://www.jultrasoundmed.org/content/24/2/215.full.pdf
We need some exact statistical analysis which establishes reproducibility of the developed algorithm and its using in practice.

The output results can be stored as images (graphs) with all computed 256 values into a CSV files. These stored files have been used for statistical analysis which was realized in MS Excel with these output values. In CSV files are stored all computed values of the area, minimal and maximal intensity and also the sum of all 256 computed values (6). All these data are necessary to analysis to assessment of reproducibility.

III. STATISTICAL ANALYSIS OF ACHIEVED RESULTS

In this chapter we will discuss results of practical measurement using by the algorithm.

From the speed of decreasing area represented of the curve we can compare that:

- in case of SN – rapid decreasing is physiological, slow decreasing is pathological (hyperechogenic SN)
- in case of RN – vice versa in comparison with SN

We need exact evaluation with some criterion to diagnose. In medical statistics is very frequently used $n$-th percentile as a statistical descriptor. We use $n$-th percentile to distinguishing pathological and physiological cases. For SN measurement we use 90$^{th}$ percentile and for RN we use 10$^{th}$ percentile as the referential border between pathology and physiology [13].

Firstly we must consider computed values of the area as the 1-D vector in the form

$$v = (a_{T0}, a_{T1}, \ldots, a_{T255})$$

where $a_{Tj}$ represents the computed area for $T = j$. Each measurement is represented by this vector $v$. Mathematically is computed $n$-th percentile as:

$$P_{n} = \left( \frac{100}{N} \right) \left( \frac{n-1}{2} \right),$$

where $N$ is number of values in the vector. We computed referential $n$-th percentile from physiological cases as the border between physiology and pathology.

A. Analysis with border between pathological and physiological cases

Firstly will be discussed results from SN measurement. In case of SN measurement the algorithm has been tested with 100 images from Philips HDI-5000, more than 200 images from ESAOTE My Lab 9 and 80 images from GE Vivid 7. We observed correlation coefficients between observers, between machines and between right and left temporal bone window.

To computing the referential percentile we use in this case 90$^{th}$ percentile as the border. We computed this percentile from 100 non-parkinsonic volunteers. We obtained from each measurement the vector $v$ which contains all 256 measured values. So, we have 100 vectors and each of them contains $a_{T0}$, $a_{T1}$ up to $a_{T255}$. Thus, we computed 90$^{th}$ percentile by means of (5) for 100 values of $a_{T0}$, 100 values of $a_{T1}$ up to 100 values of $a_{T255}$ and we will get desired vector in the following form:

$$v_{90} = (a_{90T0}, a_{90T1}, \ldots, a_{90T255}),$$

where $a_{90Ti}$ represent computed 90$^{th}$ percentile (6) from all $a_{Ti}$ values; $i \epsilon \{0, 255\}$. So, we obtained the vector $v_{90}$ as $P_{90}$ (6).

To diagnose the algorithm automatically computes difference

$$diff = v - v_{90}$$

which we used to evaluation. Subsequently is computed the sum of diff given by

$$\Sigma diff$$

More details about this methodology are available in [8], [13]. We computed separately 90th percentile for GE and ESAOTE device.

We computed separately 90th percentile for GE and ESAOTE device.

![Decreasing area of defects in mm²](image)

**Fig. 13** A measured patient with border as 90th percentile

We can use different border as the percentile which we computed. In other words, we can use different percentile for each structure according to computed results. We have computed percentile for SN (Fig. 13) for ESAOTE and GE separately and also proofing percentile for RN which has been verified by neurosonologist. In another cases we used the summation only to distinction of physiology and pathology.

### B. Kappa analysis

In case of SN measurement we proved reproducibility of the algorithm [8], [13] with correlation analysis and also kappa coefficients. Reproducibility has been verified by achieved kappa-coefficient for 2 observers (Cohen’s kappa). Kappa-coefficient is given by

\[
\kappa = \frac{\sum_{i=1}^{n} n_{ii} - \sum_{i=1}^{n} m_{ii}}{n - \sum_{i=1}^{n} m_{ii}}
\]

(12)

In case of \(n = 2\) is used the Cohen’s kappa, because were two raters only. Cohen’s kappa is defined by the following formula (the special case of (10) for \(n = 2\):

\[
\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}
\]

(13)

where \(Pr(a)\) is the relative observed agreement among raters and \(Pr(e)\) is the hypothetical probability of chance agreement. Kappa coefficient is confined to the range \([0, 1]\), higher is better level of agreement between observers.

### C. Achieved results of the statistical analysis in practice

Correlation analysis achieved results (see [8], [12], [13]):

- correlation between 90th percentiles, thus \(r(90p_{ES}, 90p_{GE}) = 0.995\)

We obtained results for 100 non-PD volunteers and for 30 parkinsonic cases as follows:

- in case of 100 non-PD volunteers were only 7 patients classified as pathological
- in case of 30 parkinsonic examined patients were only 2 classified as purely physiological [13]

Reproducibility analysis of kappa-coefficient (for details see [13]):

- for GE device average kappa coefficient from all left/right temporal window images is \(\kappa = 0.854\)
- for left temporal window images from GE and ESAOTE is average kappa-coefficient \(\kappa = 0.943\)

Both of these kappa coefficients are generally rated as “almost perfect” by general kappa coefficient rating.

We also investigated ROC analysis for SN which visualized reliability in form of ROC curve\(^{14}\) [8].

---

More details about reproducibility of achieved results of measurements in SN in [2], [8] and [13].

D. Raphe nuclei measurement analysis

In case of RN measurement were examined totally of 59 patients with panic disorders. In B-MODE is pathological anechogetic RN and we use \(10^{th}\) as the referential border in comparison with SN [13] but we use the sum of measured values (6), see Fig. 7, defined by:

\[
SUM = \sum_{T=0}^{255} a_T
\]

where \(a_T\) represents \(j\)-th value of the area for \(T\) [13]. Also for RN we proved reproducibility for 80 volunteers; only 7 patients have been incorrectly classified as pathological, for details in [13]. More about methods used in medical statistics in [14].

IV. EXPERIMENTS IN SCILAB AS AN OPEN-SOURCE SOLUTION

Apart from creating the application in MATLAB which is fully working, we also want to create the same application in Scilab as an open-source solution. So, price of MATLAB with needed toolboxes for image processing and compiling into executable application is higher than $1000. Scilab\(^{15}\) is free, open-source IDE for computing and scientific applications. Also similar to MATLAB, Scilab provides toolboxes for specific applications. For image processing Scilab provides 2 toolboxes namely SIP\(^{16}\) and SIVP\(^{17}\). Advantage of these toolboxes is very similar syntax of the basic commands which we use in MATLAB, e.g. imread, imshow, imroi, rgb2gray, etc. Even though we have 2 toolboxes for image processing, in Scilab we cannot create the same application as in MATLAB. Both of the image processing toolboxes have limitations and we cannot use both of them simultaneously. The following limitations have been found:

- SIVP can read DICOM, SIP not (bitmaps only)
- SIP can work with ROI as layer, SIVP does not support ROI
- In SIP we cannot modify basic elliptical ROI to needed angle
- We cannot simply use computing the area for all \(T\) in one step with plotting the graph
- SIP does not work with new Scilab versions in which we use SIVP

These limitations result in fact that we cannot create the same application. Plotting of the graph of the decreasing area also is not ideal as in MATLAB, see the following figure.

V. CONCLUSION

The presented paper is focused on medical image processing of ultrasound B-images in practice. The main goal of the presented work was to create a semi-automatic algorithm to echogenicity detection and its verification and validation in practice.

We created the algorithm based on evaluation of echogenicity in case of Parkinson’s Disease characterized by hyperechogenicity in substantia nigra and also we tested the developed algorithm for raphe nuclei area. Inseparable part of this processing is validation of input image, their global average intensity more precisely. It is necessary step for the further processing to avoidance of mistaken results. We proved that for three different ultrasound machines is average intensity about 25 and it was verified by histogram analysis. So, we ensure equal initial conditions for all images.

\(^{15}\) http://www.scilab.org/
\(^{16}\) http://siptoolbox.sourceforge.net/
\(^{17}\) http://sivp.sourceforge.net/
After developing we also analyzed this algorithm in practice by statistical analysis. This analysis proved that the developed algorithm is usable in medical practice and is reproducible. We tested it with 3 observers and with 3 different ultrasound machines and another tests are in progress. The algorithm has been verified and tested by neurologists and show that is usable in practice and could be used in different cases of analysis of B-images due to this principle of echogenicity, not only in neurosonology but also in another examinations with diagnostic ultrasound. Achieved correlation coefficients among observers \((r > 0.82)\) and machines \((r > 0.88)\), kappa-coefficients \((\kappa > 0.8)\) and ROC analysis show the developed algorithm is independent on used machine, image resolution and input format with automatic verification of image quality against mistaken detection.

In future will be tested other ROI in B-images to additional verification of reproducibility in practice.

ACKNOWLEDGMENT

We would like to thank Ass. prof. Ing. Sosík Petr, Dr., the investigator number CZ.1.05/1.1.00/02.0070 listed as OP VaVpl 1.1 program, applicants VŠB TÚ partnership with the Silesian University in Opava.

Also we would like to thank Assoc. Prof. Školoudík David, Ph.D. for his help, collecting needed data and verification of the developed algorithm in medical practice.

REFERENCES


