AN EVALUATION OF BRAIN TISSUE CLASSIFICATION IN NON-COMPENSATED ULTRASOUND IMAGES

[1]*Ewout Vansteenkiste, Bruno Huysmans, Wilfried Philips*

[1]ervsteen@telin.ugent.be, bhuysmans@telin.ugent.be
[1] Ghent University, department TELIN, St.-Pietersnieuwstraat 41, Ghent, Belgium

ABSTRACT

In this article we present new results on the classification of the neonatal "White Matter Damage" brain disease. One of the common diagnostic methods nowadays used in clinical practice is the visual inspection of Ultrasound images of the neonatal brain. Given the poor image quality of Ultrasound images and the different machine settings used in practice, this diagnosis highly depends on the interpretation of the medical doctor and is subjective to some degree. In this paper we investigate if the texture present in the images could have prognostic implications for detecting affected tissue, and thus help us in creating semi-automatic tools to assist the experts. We try not to compensate for the machine settings as was done in former experiments because this compensation is often machine dependent and quite tricky. We have to guess up to some degree what goes on inside of the Ultrasound machine. As a main contribution will show it is possible to get very high classification rates without this preprocessing which is a great step forward in the quantitative analysis of the images.

1. INTRODUCTION

The aim of this research is to assist medical doctors in making a better diagnosis of the White Matter Damage (WMD) brain disease, which occurs on 20 to 50 percent of newborns with a very low birth weight (< 1500 g) [1]. We do this by developing semi-automatic texture analysis tools as well for the classification as for the segmentation of the affected parts of the brain. In what follows we will only focus on the classification of affected and non affected tissue.
When capturing an Ultrasound image the medical expert selects various scanner settings, such as the Gain (the amplification of the received signal), the Power (the amplitude of the emitted waves) and Time Gain Compensation (using different levels of amplification for signals received from different depths), ... as to optimize the image on display. These settings differ from patient to patient and from expert to expert and influence the grey values displayed, see Fig. 2. Since we want to compare images quantitatively with respect to texture statistics directly computed from the grey values, it might be useful to normalize the images first so that they are independent of the scanner settings.

In the past, a compensation algorithm that constructs such a standard image was developed [2]. This machine model as we could call it makes some assumptions on the way the real Ultrasound machine forms the images, see Fig. 3. Here for, experimental data obtained from images with different parameter settings was needed. Although the framework fits its purpose well, it is machine dependent and thus not applicable to all images. That's why it would be nice to eliminate this step and try to work on the raw, non-compensated images, trying to find tissue texture features that are insensitive to all and different machine settings.
In [3] was shown that it was possible to detect affected tissue by using the appropriate features but that the classification is highly dependent on the compensation algorithm used. We should comment though that the images used for this research came from a small data set.
Since this former experiment we have received a bigger data set from a different, more modern, Ultrasound machine. This makes further experiments and statistical validation possible. The paper is organized as follows: In the next section the new experimental setup is described. Section 3 explains the used feature extraction methods. Section 4 reviews the technique used to reduce and classify the features. The results are discussed in section 5, followed by a conclusion in section 6.

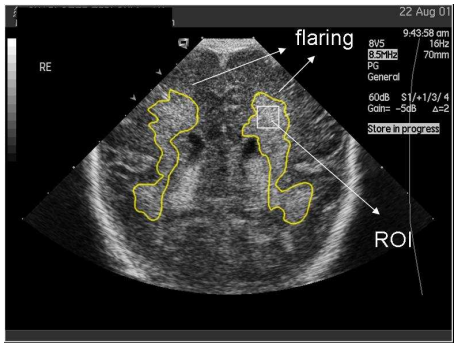

Figure 1: Symmetrical flares (affected tissue) in the ultrasound image (coronal cross-section), manually delineated by an expert. The ROI in which texture features are calculated is also shown.

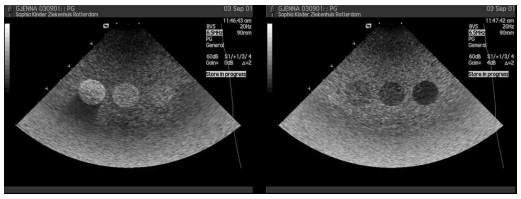

Figure 2: left image: image of a hardware phantom containing 3 cilinders captured with Gain = 4 db, right image: same hardware fanthom captured with Gain = 0 db

2. EXPERIMENTAL SETUP

In [3] only 35 images, 21 affected and 14 non affected ones, a mixture of coronal and saggital sections, captured by an Ul-

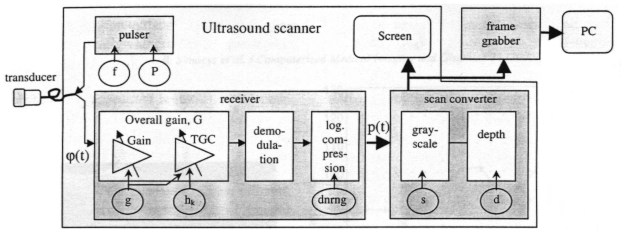

Figure 3: overview of Ultrasound machine model used in [2].

tramark Ultrasound machine, were taken into account. Given the small number of samples, multiple Regions of Interest (ROI), square regions in which the texture features are calculated - see Fig. 1, were selected per image. In that way not all samples were statistically independent. Now our new data set consists of 60 images, 30 affected and 30 non affected images, again a mixture of coronal and saggital sections, captured on an Acuson Sequoia 512 Ultrasound machine. Important to mention is that the images were taken by the same medical doctor as the first ones and no ROI had to be selected double, implying better generalization properties.
Yet some drawbacks still remain. As mentioned in the introduction the compensation algorithm as it was developed is not machine independent, that's why we do not (want to) use it on the new data set. Further more the angle of incidence of the ultrasound waves varies over the images. On some images the ventricle of the brain is clearly visible, while on other images we just see the tissue in front of it. This could be disadvantageous because the observed texture may depend on this angle of incidence (this is still investigated at the moment). On the other hand, mixing the angles of incidence and even stages of the disease makes the system more useful, because a bigger variety of images can be used as input.

3. FEATURE EXTRACTION

Here 5 texture feature sets were computed from the manually chosen ROI. These features describe the spatial interrelationships, arrangement of the image pixels, and are commonly used in many medical pattern recognition tasks [4],[5]. As you will see we choose to work only on texture features extracted from the image domain. We could also have computed Gabor or Wavelet based features for instance, though since the application should run in real time, these techniques are more complicated and less suitable in clinical practice. If the results we obtain with the simpler techniques should not perform well enough we can still incorporate those.

A) Gray Level Co-occurrence Features. Haralicks' co-occurrence matrix [6] is intuitively a 2-dimensional histogram of the grey values of pixel pairs located at a predefined distance d under a specific angle in the intensity image. In our case we made the matrix independent of by averaging out over 8 predefined angles $= \{\frac{2k}{8} | k = 1..8\}$. Let I be an $M \times N$ (intensity) image, $\triangle x = d\cos$ and $\triangle y = d\sin$, then the entry on position (i,j) in the matrix is given by

$$P_d(i,j) = \frac{1}{R} \sum_{m=\triangle x}^{M-\triangle x} \sum_{n=\triangle y}^{N-\triangle y} \in (f(m,n) = i \wedge f(m+\triangle x, n+\triangle y) = j),$$

with $R = \sum_{m=\triangle x}^{M-\triangle x} \sum_{n=\triangle y}^{N-\triangle y} \in ((m,n) \in I)$ is a normalization factor, (x) the Kronecker delta function, and $f(m,n)$ the grey value of pixel (m,n). From this matrix multiple first and second order parameters were calculated:
1) Mean gray level 2) Variance of gray level 3) SNR 4) Angular second moment 5) Contrast 6) Correlation 7) Sum of squares: variance 8) Inverse difference moment 9) Entropy 10) SNR 11) Kappa 12) Co-occurrence mean

B) Sum and Difference Histograms. Unser [7] developed a technique to extract features from the histograms of both sums and differences between pairs of grey values separated by a distance d in a direction , as a quicker alternative to the co-occurrence matrix. Let y_1 and y_2 denote 2 pixels separate by the distance vector $\mathbf{d} = (d_1, d_2)$:

$$\begin{cases} y_1 = y_{k,l} \\ y_2 = y_{k+d_1,l+d_2}, \end{cases}$$

then the sum and difference histograms are calculated from the sums $s_{k,l}$ and differences $d_{k,l}$,

$$\begin{cases} s_{k,l} = y_{k,l} + y_{k+d_1,l+d_2} \\ d_{k,l} = y_{k,l} - y_{k+d_1,l+d_2} \end{cases}$$

the sums $s_{k,l}$ take on values in the interval $[0, 2G]$, the differences in the interval $[-G, G]$, G denoting the maximum grey value in the image. The sum $P_S(i)$ and difference histogram $P_D(i)$ are then defined as:

$$\begin{aligned} P_s(i) &= h_s(i)/N; \quad (i = 0, \ldots, 2G) \\ P_d(j) &= h_d(j)/N; \quad (j = -G, \ldots, G) \end{aligned}$$

$$\text{with} \begin{cases} h_s(i; d_1, d_2) = h_s(i) = \#\{(k,l) \in ROI, s_{k,l} = i\} \\ h_d(j; d_1, d_2) = h_d(j) = \#\{(k,l) \in ROI, d_{k,l} = j\} \\ N = \sum_{i=0}^{2G} h_s(i) = \sum_{j=-G}^{G} h_d(j) \end{cases}$$

Concerning the choice of , the conclusions made in A) stay valid. The four extracted features are: 1) Mean 2) Angular second moment 3) Contrast/Variance 4) Entropy.

C) Statistical Features. Amelung developed a system AST [8] to compute features derived from the grey level and gradient histograms. He defines the 2 gradient histograms as the image histograms after convolution with the Sobel filter masks. Each histogram is used to compute 6 features:
1) Mean 2) Variance 3) Third moment 4) Fourth moment 5) Angular second moment 6) Entropy.

D) Run length Matrix. This method assumes lengths of runs in different directions can serve as a texture description. A *'run'* is a set of pixels of constant intensity on a line, under a given orientation. The run length matrix is obtained by counting the number of runs of a given length for each

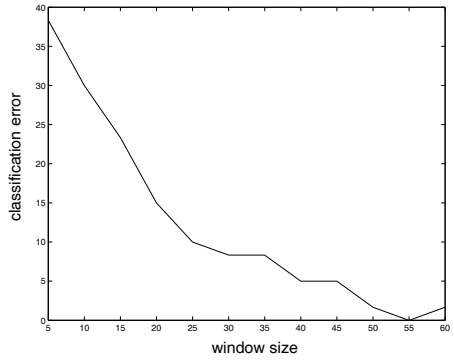

Figure 4: The classification error rate in function of the window size of the square ROI

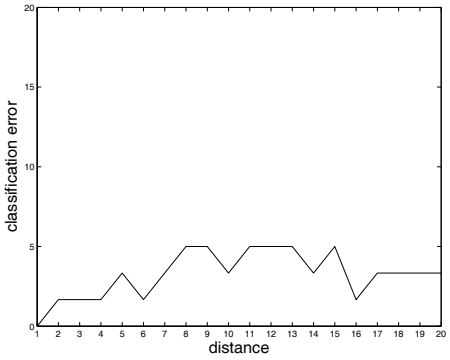

Figure 5: The classification error in function of the co-occurrence distance d for window size 55×55 pixels

grey level. Let P denote the run length matrix for an angle , then:

$$P\ (g,d) = a_{g,d}$$

Where $a_{g,d}$ stands for the number of runs of connected pixels of length d in the direction of all of which have the grey value g. Before computing the run length matrix, the images were sent through a low pass filter to reduce the noise and the grey levels were coarsely quantized to get sufficiently high run lengths. Best results were obtained by reducing to 8 gray levels using histogram equalization. Concerning the choice of , the conclusions made in A) again stay valid. 11 features are then extracted [9].
1) Short run emphasis 2) Long run emphasis 3) Gray level distribution 4) Run length distribution 5) Run percentage 6) Low gray level emphasis 7) High gray level emphasis 8) Long run high gray level emphasis 9) Long run low gray level emphasis 10) Short run high gray level emphasis 11) Short run low gray level emphasis.

E) Laws' Texture Energy Measures. Laws' texture measures are computed by first applying small convolution kernels to the image, and then combining statistics (e.g. energy) of the resulting images to extract texture features. The 2-D convolution kernels typically used for texture discrimination are generated from the following set of five one-dimensional convolution kernels of length five:

$$\begin{aligned} L &= (1,4,6,4,1) \\ E &= (-1,-2,0,2,1) \\ S &= (-1,0,2,0,-1) \\ W &= (-1,2,0,-2,1) \\ R &= (1,-4,6,-4,1) \end{aligned}$$

where L performs local averaging, E is an edge detector, S detects spots and the W and R vectors act as wave detectors. From these one-dimensional convolution kernels, we can generate 25 different two-dimensional convolution kernels by convolving a vertical 1-D kernel with a horizontal 1-D kernel. We used the texture energy of the filtered images to extract 14 texture features [10].

4. CLASSIFICATION

Computing all these features and combining them we end up with huge numbers (> 150 features). Given the number of samples $N = 60$ we used a maximum of $l = 3$ features so that the ratio $\frac{N}{l} = 20$ is sufficiently high. This to overcome the curse of dimensionality and in order to have good generalization properties this ratio should at least be 20 according to [11].
We did not reduce our feature space by any PCA search but by a simple, though computationally extensive Sequential Forward Search up to 3 features. Following the feature space reduction, the (hard) classification of the brain tissue into affected or non affected was done using a MAP Bayesian classifier with (multi)normal class distributions. Because of the size of the data set we used the same data for both purposes applying the leave-one-out principle. The error rate of the classification is computed as:

$$\text{Error rate } [\%] = 100 \times \frac{\#\text{ misclassified samples}}{\#\text{ samples}}$$

5. DISCUSSION

The aim of this study was to compare different texture feature sets. The ability of discriminating between affected and non affected tissue without having to compensate the images first was the most important issue.
Table 1 shows us that all but one of the texture feature extractors perform significantly better on the new data set, without compensation. In the case of the co-occurrence matrix we even achieve a perfect classification, which was never possible in the former data set even with compensation. This is a very promising result for the application in medical practice. As mentioned before, the bigger sample size (about doubled) of the our new data set makes these results also more statistically relevant then the former ones. A simple t-test suffices to prove this.
Since the co-occurrence features outperformed the others we will now focus a bit more on them. We tested different window sizes for the ROI ranging from 5×5 up to 60×60 pixels. Fig. 4 shows that we reach a perfect classification for a window size of 55×55 pixels, this is also the window size for which we obtain optimal results for the other techniques. Thus we can conclude that for this particular problem this is optimal. We also did some further experiments on the co-occurrence distance d. We tested distances ranging from 1 to 20, which is the an upper bound to keep a significant amount of entries in the matrix. Fig. 5 shows us that $d = 1$ gives

Lowest Error rate [%]	New data set nc	Former data set nc	Former data set c
Co-occurrence matrix	**0**	9	3
Sum and Difference histograms	6.6	2	11
Statistical features	3.3	6	1
Run length matrix	**0.5**	17	16
Laws' texture energy measures	25	29	29

Table 1: Comparison of the classification results with and without compensation algorithm obtained with each of the 6 tested feature sets. As well with our new as with our old data set. nc = non-compensated, c = compensated

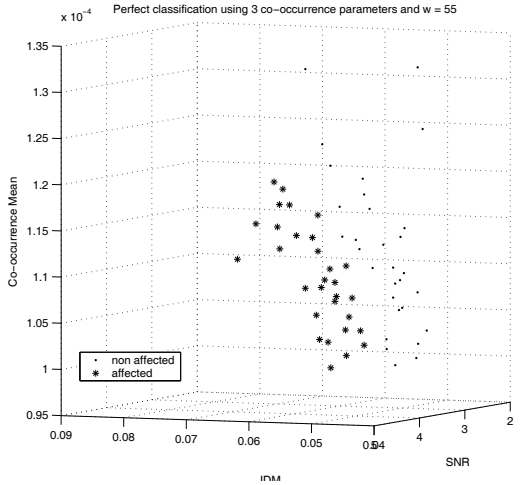

Figure 6: Classification of the samples according to the co-occurrence's Inverse Difference moment, Signal to Noise Ratio and Co-occurrence mean

us the best results. Similar results are obtained for the Run Length Matrix and Sum and Difference histograms, concerning this distance.
As for which parameters actually led to the best classification, we noticed that the Inverse Difference Moment, Co-occurrence mean and the Signal to Noise ratio, came up as best parameters, see Fig. 6.
Since the first two are related to the contrast present in the image we might explain this by the fact that they are not (as much) affected by the machine setting, since although the power and gain may brighten or darken the overall image, the contrast is less affected. Also, since we are more or less looking at the same depth in each image, the time gain compensation, which is the hardest to simulate, is of minor importance here.

6. CONCLUSION

We found relevant means to classify affected from non affected brain tissue without having to compensate for the machine settings first. 3 of the co-occurrence based feature seem to be independent of the machine settings. With the compensation already good results were obtained yet we have can conclude now that the texture features we found, outperform those in the former research.
Further validation is still necessary, for sure on two fields. First of all, when new data sets become available, which unfortunately takes time given the nature of the patients we work on, we should try and separate the test and training set even more. Secondly a more empirical model for the class distributions in the Bayesian classifier has to be taken into account.

REFERENCES

[1] A. Peelen and P. Govaert, *Chorioamnionitis and flaring*, Sophia Children's Hospital, Rotterdam, Holland, 2002.

[2] B. Simaeys, W. Philips, I. Lemahieu, and P. Govaert, "Quantitative analysis of the neonatal brain by ultrasound," *Computerized Medical Imaging and Graphics*, vol. 24, pp. 11–18, 2000.

[3] B. Huysmans, E. Vansteenkiste, P. Govaert, and W. Philips, "An evaluation of texture classifiers for the detection of periventricular leukomalacia," *Proceedings of the IEE Medical Signal and Information Processing Conference - MEDSIP 2004, Sliema, Malta*, pp. 201–206, 2004.

[4] Y. Kadah, A. Fara, J. Zurada, A. Badawi, and A. Youssef, "Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 466–478, 1996.

[5] O. Basset, Z. Zun, J. Mestas, and G. Giminez, "Textural analysis of ultrasonic images of the prostate by means of co-occurrence matrices," *Ultrasonic Imaging*, , no. 15, pp. 218–237, 1993.

[6] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, November 1973.

[7] M. Unser, "Sum and difference histograms for texture analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 118–125, Januari 1986.

[8] J. Amelung, "Automatische bildverarbeitung für die qualitätssicherung," Dissertation, Technische Hochschule Darmstadt, Darmstädter Dissertationen D17, 1995.

[9] M.M. Galloway, "Texture analysis using gray level run lenghts," *Computer Graphics and Image Processing*, vol. 4, pp. 172–179, 1975.

[10] K. Laws, "Rapid texture identification," in *Proceedings of SPIE Image Processing for Missile Guidance*, 1980, vol. 238, pp. 376–380.

[11] A. Jain and M. Tuceryan, *Handbook of Pattern Recognition and Computer Vision*, chapter Texture analysis, World Scientific Publishing Co., 1998.