CLASSIFYING AFFECTED BRAIN TISSUE IN UNCOMPENSATED ULTRASOUND IMAGES OF NEONATES

E. Vansteenkiste*, B. Huysmans* and W. Philips*

* Ghent University - Telecommunication and Information Processing Department,
St-Pietersnieuwstraat 41, Ghent, Belgium

ewout.vansteenkiste@ugent.be

Abstract: One of the common diagnostic methods nowadays used in neonatal clinical practice is the visual inspection of Ultrasound (US) brain images of the newborns. Given the inherent poor image quality of US images, due to the speckle noise and the multiple machine settings used in practice, this diagnosis highly depends on the interpretation of the medical doctor and is subjective to some degree. We now investigate how the tissue texture as displayed in the 2D US images helps us in creating tools to assist the experts in more objective diagnoses. Moreover we present a (semi-)automatic classification of the neonatal ”White Matter Damage” brain disease. New is that 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. As a main contribution will show it is possible to get very high classification rates using co-occurrence matrix based texture features on the uncompensated images. As a validation we cross-correlate our results to a segmentation algorithm we recently developed.

1 Introduction

The main aim of this research is to assist medical doctors in making a more objective diagnosis of the White Matter Damage (WMD) brain disease, see figure 1, 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 mainly focus on the classification of affected and non-affected tissue and use a segmentation scheme merely as a validation.

When capturing an US 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 visual quality of the image on display. These settings differ from patient to patient and from expert to expert and influence the grey values displayed, see figure 2. Since we want to compare images quantitatively with respect to texture statistics directly computed from the grey values, we normally compensate the images first to some kind of reference image, so that it becomes independent of the scanner settings.

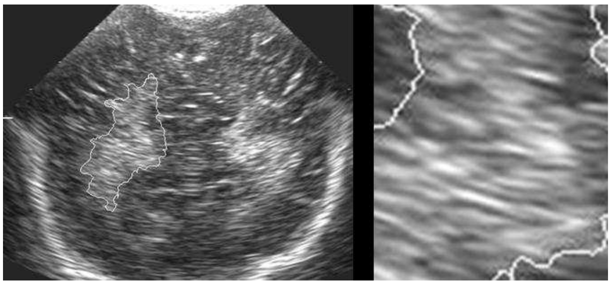

Figure 1: On the left, Flaring (affected tissue) in the US image (coronal cross-section) delineated by a white border. On the right we zoom in on the flaring, to visualize the area over which the texture features are computed.

In the past, a compensation algorithm that constructs such a reference image for our US Advanced Technology Laboratories (ATL) Ultramark 4 ultrasound machine was developed [2]. This machine model as we could call it makes some assumptions on the way the real US machine forms the images, see figure 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 images coming from newer machinery.

That’s why it would be nice to eliminate this step and try to work on the raw, uncompensated images, trying to find tissue texture features that are insensitive to all and different machine settings.

In [3] was shown that it is possible to detect affected tissue by using the appropriate features but that the classification is still highly dependent on the compensation algorithm used. Since this former experiment we have received a bigger data set from a more modern, Acuson Sequoia 512 US machine. This makes further experiments and statistical validation possible and gives us the opportunity tot test the machine dependency.

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 techniques used to reduce and classify the features. The results are discussed in section 5, followed by a cross-validaton in section 6. Finally, our conclusions and future work are discussed in section 7.

ISSN: 1727-1983 © 2005 IFMBE

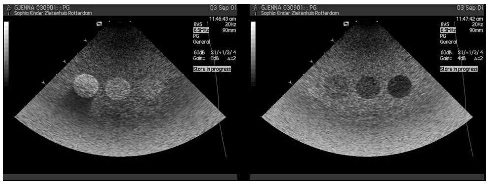

Figure 2: left image: image of a hardware phantom containing 3 cilinders captured with Gain = 4 db, right image: same hardware fanthom captured with Gain = 0 db.

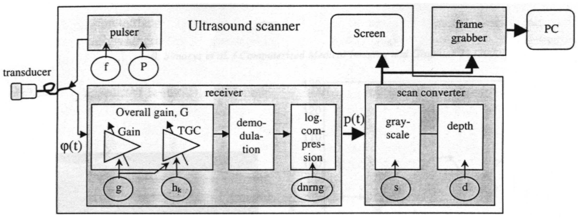

Figure 3: overview of Ultrasound machine model used in [2].

2 Experimental setup

In [3] only 35 images, 21 affected and 14 non-affected, a mixture of coronal and saggital sections captured by the ATL Ultramark 4 Ultrasound machine, were taken into account. Given the small number of samples, multiple Regions of Interest (ROI), square regions in which texture features are calculated, were selected per image. In that way not all samples were statistically independent.

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 the 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.

Although the size of the data set has almost doubled some drawbacks still remain concerning its constitution. The data set consists of images taken at different stages of the disease, ranging from newborns of the age of 1 up to 3 weeks. As it is known the flaring varies over time, it might be of importance to reclassify afterwards according to the time at which they were acquired.

Next to that the angle under which the US images is taken is not fixed and may vary around the 45 degrees angle range, see figure 4. Since ”White Matter Damage” spreads around the entire ventricle, one has little knowledge on the shape of the texture one might expect at different angles. Although these are drawbacks in the medical sense, the advantage is that until we receive bigger and more diverse data sets, a good classifier will cope with all these varieties which makes it more general.

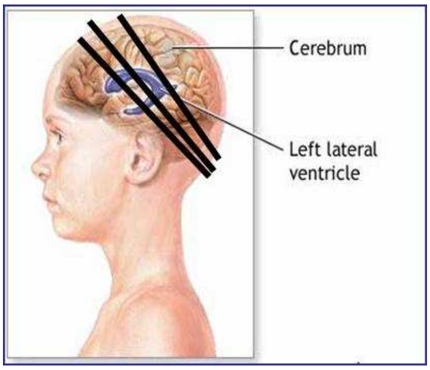

Figure 4: Our data set consists of images taken under different scanning angles. The black lines correspond to different scanner probe positions.

3 Feature Extraction

Here 5 texture feature sets were computed from the manually chosen ROI. These features describe the spatial relationships, the 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 also computed Gabor or Wavelet based features for instance but will not discuss those here. Since the application should run in real time, these techniques are more complicated and less suitable in clinical practice.

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 θ in the intensity image. In our case we made the matrix independent of θ by averaging out over 8 predefined angles $\Theta = \{\frac{2k\pi}{8} | k = 1..8\}$.

Let I be an $M \times N$ (intensity) image, $\triangle x = d\cos\theta$ and $\triangle y = d\sin\theta$, then the entry on position (i,j) in the matrix is given by

$$\begin{aligned} P_d(i,j) &= \frac{1}{R}\sum_{m=\triangle x}^{M-\triangle x}\sum_{n=\triangle y}^{N-\triangle y}\sum_{\theta\in\Theta}\delta(f(m,n)=i \\ &\quad \wedge f(m+\triangle x, n+\triangle y)=j), \end{aligned}$$

with $R = \sum_{m=\triangle x}^{M-\triangle x}\sum_{n=\triangle y}^{N-\triangle y}\sum_{\theta\in\Theta}\delta((m,n)\in I)$ is a normalization factor, $\delta(x)$ 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 θ , 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:

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

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

Concerning the choice of θ, 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 θ 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 grey level. Let P_θ denote the run length matrix for an angle θ, then:

$$P_\theta(g, d) = a_{g,d}$$

Where $a_{g,d}$ stands for the number of runs of connected pixels of length d in the direction of θ 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 θ, 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:

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

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. The Bayesian classifier is a supervised classifier where the(multi)normal pdf $P(\mathbf{x}|C_i)$ of feature set $\mathbf{x}$ belonging to class C_i, is estimated from the training data. Using Bayes' rule, the pdf $P(\mathbf{x}|C_i)$ and the *a priori* probability $P(C_i)$ that a sample belongs to a certain class are combined to calculate the *a posteriori* probability $P(C_i|\mathbf{x})$:

$$P(C_i|\mathbf{x}) = \frac{P(C_i)P(\mathbf{x}|C_i)}{\sum_{j=1}^{k} P(C_j)P(\mathbf{x}|C_j)} \qquad (1)$$

For the a priori probability $P(C_i)$, both were set to 0.5 since here we don‘t want to make any assumptions on the prevalence of the disease. Finally, a sample $\mathbf{x}$ is assigned to the class with the *maximum a posteriori* probability $P(C_i|\mathbf{x})$. 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 ability to discriminate between affected and non-affected brain tissue without having to compensate the images first was the most important issue of this research. 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. we found 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 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. Here we found that $d = 1$ gives 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 found that the Inverse Difference Moment, Co-occurrence mean and the Signal to Noise ratio, came up as best parameters, see figure 5.

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 Validation in Segmentation

In [12] we developed a segmentation scheme for the delineation of the white flaring, based on mathematical morphology. The pixel surface obtained after segmentation can also be used as an indicator for affected and

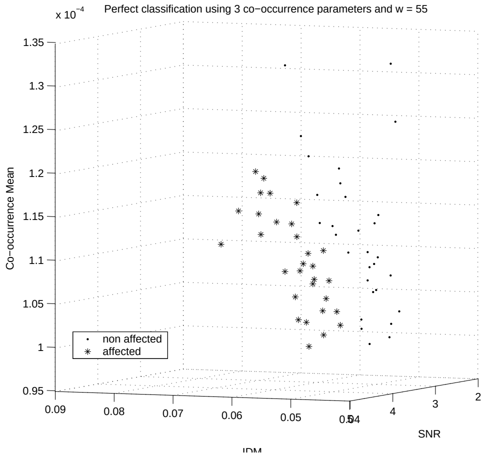

Figure 5: Classification of the samples according to the co-occurrence‘s Inverse Difference moment, Signal to Noise Ratio and Co-occurrence mean

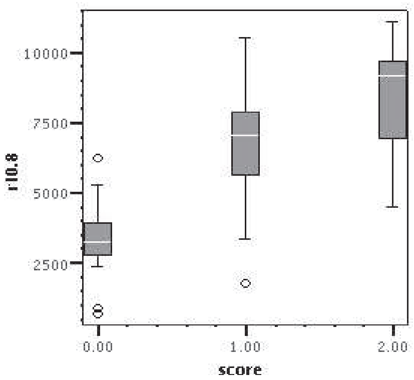

Figure 6: the boxplot of scoring grade on the X-axis versus combined pixel number of the flaring area on the Y-axis

non-affected tissue. Blinded to classification 40 US images were segmented by a clinician. The images were coronal sections through the atrium, all taken with the Acuson Sequoia scanhead, at identical gain and depth. All ultrasound frames were of preterms $\leq$ 32 weeks gestation who had first week MRI or where lesions became cystic. The patients were classified as follows : [0 normal MRI], [1 haemorrhagic WMD on MRI], [2 extensive cystic WMD]. As one can see, here the classification was done into 3 classes, there where we take class 2 and 3 as being one and the same. In figure 6 the boxplot of scoring grade on the X-axis versus combined pixel number of the flaring area on the Y-axis is shown. Comparing the normal group with the pathological one (1 and 2) the two-tailed P value is < 0.0001, considered extremely significant using the Mann-Whitney U test. Normal flaring extent was 3383 pixels (sd 1390). Pathological flares exceeded + 2sd of normal in 17 of 22 cases.

When we compare our classification results based on the co-occurrence matrix to the one based on the the pixel

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

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

surface of the segmentation we now end up with the misclassification of 6 out of the 40 samples (15%). This means that either the classifier up to now is still overtrained on the data set either the accuracy of the segmentation technique is not 100%. Here again bigger and more diverse data sets should point which of both assumptions holds.

7 Conclusion and Future Work

We succeeded in perfectly classifying the new data set using 3 of the co-occurrence based texture features, without compensating for the machine settings used. A distance $d = 1$ and window size of 55*x*55 pixels for the ROI used appeared optimal. When comparing to the classification through our segmentation scheme we see that 85% of our data set is classified in the same way. Overall we can conclude this is a good step towards more objective (semi-)automatic tools that could be embedded in bedside diagnosis equipment. As was mentioned before already, even bigger and more diverse data sets should lead to even better generalization properties of our results up to now. Staging information comes into play as well as the angle-dependency of the texture. The cross-validation using the segmentation technique is also still under investigation as well as the co-registration with 3D MRI images.

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*, 24:11–18, 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*, pages 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*, 15(4):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*, (15):218–237, 1993.

[6] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, November 1973.

[7] M. Unser. Sum and difference histograms for texture analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1):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*, 4:172–179, 1975.

[10] K. Laws. Rapid texture identification. In *Proceedings of SPIE Image Processing for Missile Guidance*, volume 238, pages 376–380, 1980.

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

[12] E. Vansteenkiste, A. Pizurica, and W. Philips. Improved segmentation of ultrasound brain tissue incorporating expert evaluation. *Proceedings of the IEEE Engeineering in Medicine and Biology Society - EMBC 2005, Shanghai, China*, 2005.

ISSN: 1727-1983 © 2005 IFMBE