Abstract—Image quality assessment plays an important role in various image-processing applications. In this work we evaluate image quality criteria based on mean square error, signal to noise ratio, and quality and structural similarity indices on ultrasound imaging of the carotid artery. These criteria as well as statistical and texture features were computed on 80 images recorded from two different ultrasound scanners before and after despeckle filtering, and after despeckle filtering and normalization. Results showed that image quality was improved after despeckle filtering and normalization for both scanners. This finding is also in agreement with the optical perception evaluation carried out by two vascular experts.

I. INTRODUCTION

Ultrasound imaging is a powerful non-invasive diagnostic tool in medicine [1] [2]. However, like all imaging modalities, ultrasound is subject to a number of artifacts that degrade image quality and compromise diagnostic confidence [1]. For medical images, quality can be objectively defined in terms of performance in clinically relevant tasks such as lesion detection and classification [3]. Most studies today have assessed the equipment performance by testing diagnostic performance of multiple observers, which also suffers from intra- and inter-observer variability. Although this is the most important method of assessing the results of image degradation, few studies have attempted to perform physical measurements of degradation [4]. The aim is to measure the amount of degradation within a system, using a number of methods, in order to locate the most useful indicators of loss of image quality. A good quality index for the comparison of two images should be: a) extracted form structural information, b) calculated on a small region such that small variations can be detected, and c) based on the regional gray-level distribution.

The major performance-limiting factor in visual lesion detection in ultrasound imaging is a multiplicative noise called speckle, that makes the signal or lesion difficult to detect [2][4]. In a recent study [2], we have shown that de-speckle filtering improves the optical perception of the physician in the assessment of ultrasound imaging of the carotid artery. The objective of this study is to evaluate the image quality of ultrasound imaging of the carotid artery on two different ultrasound scanners before and after de-speckle filtering, and after despeckle filtering and image normalization, by statistical and texture analysis, by optical perception evaluation, as well as by image quality metrics.

In the following Section, theoretical concepts on image quality are introduced. In Section III. A and Section III. B, the two different ultrasound scanners, and the de-speckle filter used are presented respectively. In Section III. C image normalization is presented, whereas in Section III.-D and Section III. E, statistical and texture analysis and the evaluation metrics are presented respectively. Sections IV and V present the results and concluding remarks.

II. IMAGE QUALITY

In order to be able to design accurate and reliable quality metrics, it is necessary to understand what quality means to the observer. An observer’s satisfaction when watching an image depends on many factors. One of the most important is of course image content and material. Research made in the area of image quality showed that this depends on many parameters, such as: viewing distance, display size, resolution, brightness, contrast, sharpness, colorfulness, naturalness, and other factors [3].

It is also important to note that there is often a difference between fidelity (the accurate reproduction of the original on the display) and perceived quality. Sharp images with high contrast are usually more appealing to the average observer. Likewise, subjects prefer slightly more colorful and saturated images despite realizing that they look somewhat unnatural [6]. For studying visual quality some of the definitions above should be related to the human-visual system. Unfortunately, subjective quality may not be described by an exact figure, due to its inherent subjectivity, it can only be described statistically. Even in psychological threshold experiments, where the task of the observer is to give a yes or no answer, there exists a significant variation between observer’s contrast sensitivity functions and other critical low-level visual parameters. When speckle noise is apparent in the image, the observer’s differing experiences with noise are bound to lead to different weightings of the artifact [5]. Researchers showed that experts and non-experts, with respect to image quality, examine different critical image characteristics to form their final opinion [7].

In the light of the above difficulties, testing procedures for subjective quality assessment are discussed in the next Section.
III. METHODOLOGY

A. Ultrasound Imaging Scanners

Two different ultrasound scanners were used in our study, the ATL HDI-3000 and the ATL HDI-5000.

The ATL HDI-3000 ultrasound scanner is equipped with 64 elements fine pitch high-resolution, 38 mm broadband array, a multi element ultrasound scan head with an operating frequency range of 4-7 MHz, an acoustic aperture of 10x8 mm and a transmission focal range of 0.8-11 cm.

The ATL HDI-5000 ultrasound scanner is equipped with 256-element fine pitch high-resolution 50 mm linear array, a multi element ultrasound scan head with an extended operating frequency range of 5-12 MHz and real spatial compound imaging. The scanner increases the image clarity using SonocT imaging by enhancing the resolution and borders, and interface margins are better displayed. Several tests made by the manufacturer showed that, the ATL HDI-5000 scanner was overall superior to conventional 2D imaging, primarily because of the reduction of speckle, contrast resolution, tissue differentiation, and the image was optically better.

In this study 80 ultrasound images of the carotid artery were acquired from each scanner. The images were recorded digitally on a magnetic optical drive with a resolution of 768x756 pixels with 256 gray levels.

B. Despeckling

The working principle of the de-speckle filter despeck
used in this study, which was applied six times iteratively on each image, may be described by a weighted average calculation using sub region statistics to estimate statistical measures over pixel windows (typically 5x5, 7x7, or 9x9 sliding pixel windows). It assumes that the speckle noise model has the following multiplicative form 

\[ g_{ij} = f_{ij} n_{ij}, \quad i,j \in N \]  

(1)

where \( g_{ij} \) represents the noise pixel in the middle of the moving window, \( f_{ij} \) represents the noise-free pixel and \( n_{ij} \) is a Rayleigh distributed noise on pixel. Hence the algorithms in this class may be traced back to the following equations 

\[ f_{ij} = \bar{g}_{ij} + k_{ij} (g_{ij} - \bar{g}_{ij}) \]  

(2)

where \( f_{ij} \) is the new estimated pixel, \( g_{ij} \) is the old pixel in the middle of a moving window, \( \bar{g}_{ij} \) is the local mean value of an MxN region, \( k_{ij} \) is a weighting factor with \( k \in [0,1] \), and \( i,j \) the absolute pixel coordinates. The factor \( k_{ij} \) is a function of the local statistics in a moving window. It is found in the literature and is derived as 

\[ k_{ij} = \frac{\sigma^2_{g}}{\sigma^2_{g} + \sigma^2_{n}} \]  

(3)

with \( \sigma^2_{g} \) and \( \sigma^2_{n} \) the variances of the original image and noise respectively.

C. Image Normalization

The images were normalized manually by linearly adjusting the image so that the median gray level value of the blood was 15-20, and the median gray level of the adventitia (artery wall) was 180-200. The scale of the gray level of the images ranged from 0-255. This normalization using blood and adventitia as reference points was necessary in order to extract comparable measurements in case of processing images obtained by different operators or different equipment.

D. Statistical and Texture Analysis

Texture contains important information, which is used by humans for the interpretation and analysis of many types of images. It may provide useful information about object characterization in ultrasound images. The following statistical and texture features (SF) were extracted from the original, and the processed images to evaluate quantitatively the performance of each scanner:

a) Statistical Features (SF): 1) Mean, 2) Median, 3) Standard Deviation, 4) Skewness, 5) Kurtosis and 6) Speckle index (Mean/Standard Deviation), b) Spatial Gray Level Dependence Matrix-range values (SGLDM) selected features as proposed by Haralick et al. 

1) Entropy, 2) Contrast, and 3) Angular Second Moment.

E. Image Quality and Evaluation Metrics

Differences between images were evaluated using the following image quality and evaluation metrics, which were used as statistical measures, between the original \( (g_{ij}) \) and the processed \( (f_{ij}) \) images. The following measures, which are easy to compute and have clear physical meaning, were computed:

a) The mean square error

\[ \text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (f_{ij} - g_{ij})^2 \]  

(4)

which measures the quality change between the original and processed image.

b) The randomized mean square error, which is the square root of the squared error averaged over the MxN array

\[ \text{RMSE} = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (g_{ij} - f_{ij})^2} \]  

(5)

c) The error summation in the form of the Minkowski metric, which is the norm of the dissimilarity between two images, as follows:

\[ \text{Err} = \left( \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |g_{ij} - f_{ij}|^\beta \right)^{1/\beta} \]  

(6)

with \( 1 < \beta < 4 \). For \( \beta = 2 \), one obtains the RMSE expression in (5), whereas for \( \beta = 1 \), the absolute difference, and for \( \beta = \infty \) the maximum difference measure.

d) The geometric average error (GAE) is a measure, which shows if the transformed image is very bad and it is used to replace or complete RMSE. It is positive only if every pixel value is different between the original and the transformed image. The GAE is zero, if there is a very good transformation between the original and the transformed image, and high if the transformation with the original is extremely bad. This measure is also used for tele-ultrasound, when transmitting US images and is defined as

\[ \text{GAE} = \left( \prod_{i=1}^{N} \prod_{j=1}^{M} (g_{ij} - f_{ij})^{1/(\lambda \cdot \lambda)} \right). \]  

(7)
e) The signal-to-noise ratio [16]

$$SNR = 10 \log_{10} \left( \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (g_{i,j}^2 + f_{i,j}^2)}{\sum_{i=1}^{M} \sum_{j=1}^{N} (g_{i,j} - f_{i,j})^2} \right).$$  \hspace{1cm} (8)

f) The peak signal to noise ratio [16]

$$PSNR = -10 \log_{10} \left( \frac{MSE}{s^2} \right)$$ \hspace{1cm} (9)

where $s$ is the maximum pixel value in the image. The $PSNR$ is higher for a better-transformed image and lower for a poorly transformed image. It measures image fidelity that is how closely an image (despeckled) resembles usually the corrupted original.

g) The mathematically defined universal quality index [41] models any distortion as a combination of three different factors, which are: 1) loss of correlation, 2) luminance distortion, and 3) contrast distortion. Its highest value is 1 for $g_{i,j} = f_{i,j}$, while the lowest value is -1 for $f_{i,j} = 2g_{i,j} - g_{i,j}$ and is defined as

$$Q = \frac{\sigma_{fg}^2}{\sigma_f \sigma_g} \left( f^2 + \frac{\sigma_f}{\sigma_g} f^2 \right) \left( g^2 + \frac{\sigma_g}{\sigma_f} g^2 \right) -1 < Q < 1 \hspace{1cm} (10)$$

where $\sigma_{fg}$ represents the covariance between images and $\sigma_{f}$, $\sigma_{g}$ represents the standard deviations of the pixel values inside the window. It is calculated for a window size 8x8 using a sliding window without overlapping.

h) The structural similarity index between two images [14], which is a generalization of (10)

$$SSIM = \left( \frac{2\mu_f \mu_g + c_1}{\mu_f^2 + \mu_g^2 + c_1} \right) \left( \frac{2\sigma_{fg} + c_2}{\sigma_f^2 + \sigma_g^2 + c_2} \right) \left( \frac{2\mu_f \mu_g + c_3}{\mu_f^2 + \mu_g^2 + c_3} \right) -1 < Q < 1 \hspace{1cm} (11)$$

i) The optical perception evaluation, in our study, was carried out according to the ITU-R recommendations with the Double Stimulus Continuous Quality Scale (DSCQS) procedure [15]. Two vascular experts were used, an angiologist and a neurovascular specialist. They assigned a score in the one to five scale based on subjective criteria by inspecting the area around the distal common carotid artery, 2-3 cm before the bifurcation, as the objective of the experts was to visualize the intima media complex. The experts were allowed to give equal scores to more than one image in each case.

IV. RESULTS

Figure 1 illustrates the original, despeckled, normalized, and normalized despeckled images of the two ultrasound scanners. It is shown that the images for the ATL HDI-3000 scanner have greater speckle noise compared to the ATL HDI-5000 images. Moreover the lumen borders and the intima-media thickness are more easily identified in the ATL HDI-5000 normalized and normalized despeckled images.

Table I shows the results of the statistical and texture features for the 80 images recorded from each image scanner. As shown the effect of despeckle filtering for both scanners was similar, that is the speckle index was reduced, the mean and the median were preserved, the skewness and the kurtosis were reduced, thus making the image histogram more symmetric and less flattened. The entropy was increased and the contrast was reduced. The ASM was reduced for the despeckled images for both scanners and for the normalized despeckled images for the HDI-5000 scanner. No statistically significant difference was found for all features in Table I when performing the non-parametric Wilcoxon rank sum test at a less than 0.05 between the original (O) and despeckled (D), the original and normalized (N), and the original and normalized despeckled (ND) features for both scanners.

| Scanner       | ATL HDI-3000 | ATL HDI-5000 |
|---------------|--------------|--------------|
| Images        | O  | D  | N  | ND | O  | D  | N  | ND |
| Statistical Features (SF) | | | | | | | | |
| Mean          | 22.13 | 21.78 | 26.81 | 26.46 | 22.72 | 22.35 | 27.81 | 27.46 |
| Median        | 3.07  | 4.53  | 3.56  | 5.07  | 3.75  | 5.23  | 4.59  | 6.07  |
| Stand. Deviation | 40.67 | 36.2  | 45.15 | 41.48 | 41.22 | 36.7  | 45.9  | 42.31 |
| Skewness ($\sigma^3$) | 2.88  | 2.49  | 2.23  | 2.00  | 2.84  | 2.45  | 2.17  | 1.94  |
| Kurtosis ($\sigma^4$) | 12.43 | 10.05 | 7.94  | 6.73  | 12.13 | 9.82  | 7.56  | 6.43  |
| Speckle Index  | 0.29  | 0.27  | 0.25  | 0.24  | 0.28  | 0.27  | 0.24  | 0.23  |
| SGLDM-Range Values | | | | | | | | |
| Entropy       | 0.24  | 0.34  | 0.25  | 0.34  | 0.40  | 0.48  | 0.41  | 0.48  |
| Contrast      | 667   | 309   | 664   | 303   | 618   | 302   | 595   | 287   |
| ASM           | 0.36  | 0.35  | 0.38  | 0.37  | 0.37  | 0.33  | 0.39  | 0.35  |

Table II shows the results, in percentage (%) format for the optical perception evaluation made by the two vascular experts. It is clearly shown that the highest scores were obtained for the normalized despeckled images for both scanners from both experts.

Table III tabulates the image quality evaluation metrics, presented in Section III. E, for the 80-ultrasound
images recorded from each image scanner, between the O-D, the O-N, and the O-ND images. Best values were obtained for the ATL HDI-3000 scanner for the O-D images with lower MSE, RMSE, and Err3, and higher PSNR. The GAE was 0.00 for all cases, and this can be attributed to the fact that the information between the original and the processed images, remains unchanged. Best values for Q and SSIN were obtained for the O-N images for both scanners, whereas best values for SNR were obtained for the ATL HDI-3000 scanner on the O-N images.

**TABLE II.**

Optical perception evaluation for the 80 images processed from each scanner for the original (O), despeckled (D), normalized (N), and normalized despeckled (ND). Scores are expressed in percentage format

| Ultrasound Scanner | ATL HDI-3000 | ATL HDI-5000 |
|--------------------|--------------|--------------|
| Angiologist        | O  D  N  ND  | O  D  N  ND  |
|                    | 30  43  69  72 | 26  42  59  70 |
| Neurovascular      | O  D  N  ND  | O  D  N  ND  |
| Specialist         | 41  56  54  71 | 49  53  59  72 |
| Average            | 36  50  62  72 | 38  48  59  71 |

**TABLE III.**

Image quality evaluation metrics between the original-despeckled (O-D), original-normalized (O-N) and the original-normalized despeckled (O-ND) images

| Evaluation Metrics | ATL HDI-3000 | ATL HDI-5000 |
|--------------------|--------------|--------------|
| MSE                | 71.13 183.92 248.02 | 72.58 247.4 315.32 |
| RMSE               | 8.43 11.00 14.49 | 8.52 12.82 16.00 |
| Err 3              | 14.73 15.3 20.84 | 14.82 17.68 22.49 |
| Err 4              | 20.16 19.32 26.18 | 20.23 21.91 27.84 |
| GAE                | 0.00 0.00 0.00 | 0.00 0.00 0.00 |
| SNR                | 17.42 18.37 14.03 | 17.47 16.96 13.38 |
| PSNR               | 29.61 29.60 25.58 | 29.53 28.06 24.75 |
| Q                  | 0.73 0.93 0.73 | 0.72 0.93 0.72 |
| SSIN               | 0.94 0.95 0.92 | 0.94 0.95 0.91 |

**V. CONCLUDING REMARKS**

Image quality is very important in the assessment of atherosclerosis and the evaluation of the risk of stroke. In this work we have evaluated two different ultrasound scanners on 80 ultrasound images acquired from the carotid artery before, and after despeckle filtering, after normalization, and after despeckle filtering and normalization. We have evaluated the results based on statistical and texture features, on optical evaluation by two experts as well as based on image quality metrics.

Table I shows that brighter images were recorded for the ATL HDI-5000. Moreover it is shown that the entropy that is a measure of the information content of the image is higher for the ATL HDI-5000 in the cases of the D, N, and ND images. The ASM that is a measure of the inhomogeneity of the image is lower for the ATL HDI-5000. In the cases of the D and ND images, the inhomogeneity of the image is lower for the ATL HDI-5000 scanner, whereas best values for SNR were obtained for the ATL HDI-5000 scanner on the O-N images. The indexes that is a measure of the information content of the image is higher for the ATL HDI-5000. Moreover it is shown that the inhomogeneity of the image is lower for the ATL HDI-5000 scanner, whereas best values for SNR were obtained for the ATL HDI-5000 scanner on the O-N images.

**REFERENCES**

[1] N. M. Dicken, *Diagnostic Ultrasounds*, 3rd ed., Churchill Livingstone, Inc. NY, 1991.
[2] C. Loizou, C. Christodoulou, C. S. Pattichis, R. Istepanian, M. Pantziarlis, A. Nicolaides, “Speckle reduction in ultrasound images of atherosclerotic carotid plaque”, *DSP-2002, 14th IEEE Int. Conf. on Digital Signal Processing*. July 1-3, pp. 525-528, 2002.
[3] E. Krupinski, H. Kundel, P. Judy, C. Nodine, “The medical image perception society”, key issues for image perception research”, *Radiology*, vol. 209, pp. 611-612, 1998.
[4] Z. Wang, A. Bovik, “A Universal quality index”, *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, March 2002.
[5] A. Ahumada, C. Null, “Image quality: A multidimensional problem”, *A. B. Watson, ed.*., Digital Images and Human Vision, (Bradford Press: Cambridge Mass.), pp. 141-148, 1993.
[6] E. A. Fedorovskaya, H. de Ridder, F. J. Blomnaert, “Chroma variations and perceived quality of color images and natural scenes”, *Color Research and Application*, vol. 22, no. 2, pp. 96-110, 1997.
[7] G. Deffner et al., “Evaluation of display image quality: Experts vs. non-experts”, *Society for Inf. Display Symp. Digest*, vol. 25, pp. 475-478, 1994.
[8] A. Philips Medical System Company, “Comparison of image clarity, SonoCT real-time compound imaging versus conventional 2D ultrasound imaging”, *ATL Ultrasound, Report*, 2001.
[9] J. S. Lee, Speckle analysis and smoothing of synthetic aperture radar images, “Comp. Graph. and Image Proc.”, vol. 17, pp. 24-32, 1981.
[10] T. Elatrozy, A. Nicolaides, T. Tegos, A. Zarka, et al., “The effect of B-mode ultrasonic image standardization of the echodensity of symptomatic and asymptomatic carotid bifurcation plaque”, *Int. Angiology*, vol.17, no. 3, pp. 179-186, Sept.1998.
[11] R. M. Haralick, K. Shanmugam, I. Dinstein, “Texture features for image classification”, *IEEE Trans. on Systems, Man., and Cybernetics*, vol. 3, pp. 610-621, Nov. 1973.
[12] T-J Chen, et al., “A novel image quality index using Moran I statistics”, *Ph. in Medic. and Biology*, vol. 48, pp. 131-137, 2003.
[13] R. Gonzalez, R. Woods, Digital Image Processing, Second Edition, Prentice-Hall Inc., 2002, pp. 419-420.
[14] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, “Image quality assessment: From error measurement to structural similarity”, *IEEE Trans. On Image Processing*, pp. 1-14, 2004 (to appeared).
[15] S. Winkler, “Vision models and quality metrics for image processing applications”, PhD, University of Lausanne-Switzerland, Dec. 21, 2000.
[16] D. Sakrison, “On the role of observer and a distortion measure in image transmission”, *IEEE Trans. Communication*, vol. 25, pp. 1251-1267, Nov. 1977.