

COMPARISON OF ULTRASOUND IMAGE FILTERING METHODS BY MEANS OF MULTIVARIABLE KURTOSIS

MARIUSZ NIENIEWSKI✉ AND PAWEŁ ZAJĄCZKOWSKI

Faculty of Mathematics and Informatics, University of Lodz, ul. Banacha 22, 90-238 Lodz, Poland

e-mail: mnieniew@math.uni.lodz.pl, zajaczkowski.pawel@math.uni.lodz.pl

(Received October 7, 2016; revised February 23, 2017; accepted April 14, 2017)

ABSTRACT

Comparison of the quality of despeckled US medical images is complicated because there is no image of a human body that would be free of speckles and could serve as a reference. A number of various image metrics are currently used for comparison of filtering methods; however, they do not satisfactorily represent the visual quality of images and medical expert's satisfaction with images. This paper proposes an innovative use of relative multivariate kurtosis for the evaluation of the most important edges in an image. Multivariate kurtosis allows one to introduce an order among the filtered images and can be used as one of the metrics for image quality evaluation. At present there is no method which would jointly consider individual metrics. Furthermore, these metrics are typically defined by comparing the noisy original and filtered images, which is incorrect since the noisy original cannot serve as a golden standard. In contrast to this, the proposed kurtosis is the absolute measure, which is calculated independently of any reference image and it agrees with the medical expert's satisfaction to a large extent. The paper presents a numerical procedure for calculating kurtosis and describes results of such calculations for a computer-generated noisy image, images of a general purpose phantom and a cyst phantom, as well as real-life images of thyroid and carotid artery obtained with SonixTouch ultrasound machine. 16 different methods of image despeckling are compared via kurtosis. The paper shows that visually more satisfactory despeckling results are associated with higher kurtosis, and to a certain degree kurtosis can be used as a single metric for evaluation of image quality.

Keywords: forward-backward diffusion, multivariate kurtosis, nonlinear coherent diffusion, speckle filtering, ultrasound images.

INTRODUCTION

Ultrasound (US) imaging is a common diagnostic tool used all over the world, compare the books by Schäberle (2005) and Sanches *et al.* (2012). However, the US medical images have some limitations due to the fact that they are covered with bright speckles, which mask the true edges/contours of the tissues in the image.

Image quality comparison and assessment is a comprehensive subject, compare the book by Wang and Bovik (2006) for all kinds of images, the book by Loizou and Pattichis (2015) specific to US images, the review papers by Chandler (2013) or Lin and Kuo (2011). In the following only a couple of most relevant aspects of image quality are discussed.

Certainly evaluation of the quality of edges is particularly critical. For this purpose, the edge detection figure of merit was introduced by Pratt (2007), p. 514. The trouble is that the figure of merit is simple to calculate for straight line edges but otherwise it becomes quite complicated. In fact, Pratt (2007) considers an example of a vertical edge only.

An obvious way to evaluate image quality is through subjective tests (Chandler, 2013). In these

tests, a group of human subjects are asked to judge the quality. The scores given by observers are averaged to produce the mean opinion score (MOS). However, subjective tests of medical images are quite expensive and difficult to conduct. The other approach to evaluate image quality is the objective one (Chandler, 2013). In this case one can choose among full reference (FR), reduced reference (RR), and no reference (NR) image quality evaluation.

FR quality assessment of US images would involve the use of a computer-generated image and there are some doubts whether the results can really be carried over to natural images. Examples of FR quality assessment include the structural similarity measure (SSIM; Wang *et al.*, 2004) and edge-strength-similarity-based metric (ESSIM; Zhang *et al.*, 2013).

NR quality assessment has been a subject of a very active research in recent years (Ferzli and Karam, 2009; Li *et al.*, 2016; Narvekar and Karam, 2011; Tong *et al.*, 2005; Zhu and Karam, 2014). A more or less typical example of NR quality assessment is the following (Tong *et al.*, 2005). Examples of images are prepared and two classes are composed: one of high quality images and one of low quality. These classes are defined based on the external knowledge.

Then a binary classifier is built for separation of these classes. For this purpose a quality metric is formulated based on probability of a new image being classified as belonging to high quality class. Virtanen *et al.* (2015) describe a database developed for purposes of testing NR image quality assessment algorithms. Wu *et al.* (2015) use a publicly available LIVE II database for subjective evaluation of images. It would be tempting to adapt the developed NR techniques to comparisons of medical US images. Unfortunately, there is no similar database for speckle-contaminated US images. Under these circumstances, Mateo and Fernández-Caballero (2009) propose the use of computer-simulated US images, where both the noise-free and noisy versions of the same image are obtained and filtering effects of various algorithms are compared. Rosa and Monteiro (2014) present a table comparing 17 despeckling filters via 15 different metrics calculated for computer simulated images. Thangavel *et al.* (2009) present a comparative study of removal of speckled noise based on various filters. Mittal *et al.* (2013) propose NR image quality assessment based on natural image statistics and do not use prior human opinion score.

The aim of the current paper is to provide a new metric, that is relative multivariate kurtosis, which could serve as a NR measure for comparison of image despeckling methods and which would be consistent with subjective human feeling. The innovation of this metric in comparison with those available in the literature is that it is the absolute measure, meaning that multivariate kurtosis is calculated for the filtered image independently of any other image, whereas the vast majority of the known metrics do make comparisons, by necessity with the original noisy image since the ideal noiseless image does not exist. Further innovation stems from the empirically confirmed observation that although the relative multivariate kurtosis directly deals only with the edges, it provides a numerical measure which is in a good agreement with a visual evaluation of the quality of images and with the medical expert's satisfaction. The important point is that the existing metrics are numerous and somewhat complementary; for example, Loizou (2013) and Loizou and Pattichis (2015) use 17 metrics and it is very hard, if not impossible, to summarize their individual contributions in a single number representing the image quality.

MATERIALS AND METHODS

THEORETICAL BACKGROUND

The multivariate and in particular bivariate kurtosis is a generalization of the well-known univariate

kurtosis β_2 , which is defined for a (univariate) probability density as a ratio of the fourth central moment to the square of the second central moment

$$\beta_2 = m_4/m_2^2. \quad (1)$$

The respective central moments are defined as

$$m_2 = \sigma_x^2 = E[(X - \mu_x)^2] \quad (2)$$

and

$$m_4 = E[(X - \mu_x)^4], \quad (3)$$

where μ_x is the mean value of the random variable X , σ_x^2 is a variance, and the symbol E denotes the expectation operator.

The meaning of kurtosis is nicely explained in DeCarlo (1997). For the normal distribution kurtosis $\beta_2 = 3$. For $\beta_2 > 3$ the distribution has a higher peak and heavier tails than the normal distribution. For $\beta_2 < 3$ the distribution has lower peak and lighter tails.

The original idea of the univariate kurtosis was generalized by Mardia (1970) for measuring how much a given multivariate probability density deviates from the multivariate normal distribution.

For our purposes a simpler set of equations referring to bivariate distribution is appropriate in accordance with the fact that we are dealing with 2-D images. A set of equations used for the calculation of the bivariate kurtosis is as follows (Caviedes and Oberti, 2004; Zhang *et al.*, 1999). The mean values of the X and Y variables are

$$\mu_x = E[X] = \sum_{i=1}^n x_i \sum_{j=1}^m f(x_i, y_j), \quad (4)$$

$$\mu_y = E[Y] = \sum_{j=1}^m y_j \sum_{i=1}^n f(x_i, y_j), \quad (5)$$

where $f(x_i, y_j)$ denotes the probability density. Similarly, the variances σ_x^2 and σ_y^2 are

$$\sigma_x^2 = \sum_{i=1}^n (x_i - \mu_x)^2 \sum_{j=1}^m f(x_i, y_j), \quad (6)$$

$$\sigma_y^2 = \sum_{j=1}^m (y_j - \mu_y)^2 \sum_{i=1}^n f(x_i, y_j). \quad (7)$$

The covariance between variables X and Y is given by the equation

$$\sigma_{xy} = \sum_{i=1}^m \sum_{j=1}^n (x_i - \mu_x)(y_j - \mu_y) f(x_i, y_j), \quad (8)$$

and the correlation coefficient ρ is

$$\rho = \sigma_{xy} / [\sigma_x \sigma_y]. \quad (9)$$

The marginal distributions $f(x)$ and $f(y)$ are

$$f(x) = \sum_{j=1}^m f(x, y_j), \quad f(y) = \sum_{i=1}^n f(x_i, y). \quad (10)$$

Now the auxiliary parameters $\gamma_{k,l}$ are defined with indices k, l assuming the values from 0 to 4

$$\gamma_{k,l} = \frac{\sum_{i=1}^n \sum_{j=1}^m (x_i - \mu_x)^k (y_j - \mu_y)^l f(x_i, y_j)}{\left(\sum_{i=1}^n (x_i - \mu_x)^2 f(x_i) \right)^{k/2} \left(\sum_{j=1}^m (y_j - \mu_y)^2 f(y_j) \right)^{l/2}}. \quad (11)$$

The bivariate kurtosis $\beta_{2,2}$ is calculated according to the following equation

$$\beta_{2,2} = \frac{\gamma_{4,0} + \gamma_{0,4} + 2\gamma_{2,2} + 4\rho(\rho\gamma_{2,2} - \gamma_{1,3} - \gamma_{3,1})}{(1 - \rho^2)^2}. \quad (12)$$

It is explained in Romeu and Ozturk (1993) and in the book by Timm (2002), p. 121, that the sample estimate of multivariate kurtosis converges in distribution to normal distribution with mean $\mu = p(p+2)$ and variance $\sigma^2 = 8p(p+2)/n$, where p is the dimensionality of the variable, and n is the number of points in the sample. In our case testing for normality would not give the information we need since we want to use kurtosis precisely for evaluating the quality of edges contained in an image. Obviously, filtering an image results in the change of the shape of edges and in the change of the local kurtosis, which is calculated in a frame covering the edge. We assumed frames of size 8×8 pixels, which seems to be a reasonable choice. For a given frame we calculated the 2-D Fourier transform and shifted the origin of the transform to the center of the frame. Then we found the magnitude of the transform and normalized it by dividing by the sum of all the pixel values. This normalized magnitude played the role of the probability density distribution described approximately by a 2-D Gaussian function.

SOFTWARE AND HARDWARE

All the described calculations were executed in Matlab. One notable exception worth mentioning is the selection of the corner points of the frames used for calculating local kurtosis. For each frame one corner point has to be specified, and we found that it can be speedily done using a plugin *Point Picker* in ImageJ (Thevenaz, 2016).

The program used for artificial US image generation was the Pseudo B-Mode Ultrasound Image

Simulator available at Sheet (2016). The size of the generated image was 128×129 pixels. The phantom images were taken using the general purpose high resolution and low contrast phantom model 1525 (Fig. 6) and the extra large cyst phantom model 571 (Fig. 9), as specified by Dansk (2016a;b).

The natural and phantom US images were taken by means of the SonixTouch ultrasound system from Ultrasonix Medical Corp., from its output which gives access to logarithmically compressed images which have not been subjected to any preprocessing and in fact are quite noisy. The SonixTouch is a specialized ultrasound machine that gives access to various signals otherwise inaccessible in a standard equipment.

The settings on the SonixTouch for the general purpose phantom image were: frequency 13.3 MHz, focal depth 4 cm, gain 61 %, dynamic range 75 dB, persistence 2, map 7, and the size of the image was 456×444 pixels. The setting for the cyst phantom image were: frequency 6.6 MHz, depth 6 cm, and the rest of parameters as for the general purpose phantom, with the size of the image 456×258 pixels. The settings for the thyroid image were: frequency 10 MHz, focal depth 4 cm, gain 50 %, dynamic range 75 dB, persistence 2, and map 4. The settings for the carotid artery were the same, except gain was 50 %, with the size of the thyroid and carotid artery images equal to 520×256 pixels.

A number of tests of despeckling methods were conducted using filters available from various sources. In particular, the code for the filters developed by Loizou and other researchers was taken from Loizou (2013) and Loizou and Pattichis (2015). Various aspects of these filters are also discussed in Loizou *et al.* (2005; 2006; 2012).

Consideration of the details of Loizou filters would take too much place and here only the names of the filters and their popular abbreviations are specified (compare also Fig. 2): Diffusion Filter DsFad; Linear Scaling Filters DsFca, DsFlecasort, DsFls; Geometric Filter DsFgf4d; Homomorphic Filter DsFhomo; Maximum Homogeneity over Pixel Neighborhood Filter DsFhomog; Homogeneous Mask Area Filter DsFlsminsc; First-Order Statistics Filter DsFlsmv; Median Filter DsFmedian; Speckle-Reducing Anisotropic Filter DsFsrad; Wavelet Filter DsFwaveltc; Wiener Filter DsFwiener2.

The Non-Local Means Filter, denoted DsNLM here, was obtained from Buades *et al.* (2005) and Manjon-Herrera (2016). The Nonlinear Coherent Diffusion Filter, denoted DsNCD, was taken from Nieniewski and Zajaczkowski (2014), and Forward-Backward Diffusion Filter DsFBD from Nieniewski

and Zajęczkowski (2016). In order to make possible comparisons all filters have to process a common image, and it was assumed that this image is log-compressed since this was the most common case from the point of view of the software availability.

KURTOSIS VS. MEDICAL EXPERT ASSESSMENT OF FILTERED IMAGES

In order to compare kurtosis with medical assessment of the quality of filtered images we juxtaposed two tables. The first, that is Table 1 contains results of calculating kurtosis for the image of the carotid artery adapted from Nieniewski (2014), Fig. 1(a), and filtered as shown in Fig. 12 below. The second, that is Table 2 contains ranking the filters according to the neurovascular specialist and is adapted from Loizou *et al.* (2014), Table 4. It is based on the evaluation of 100 asymptomatic and symptomatic images of a carotid artery.

A direct one-by-one comparison of the entries of both tables is impossible because some of their filters are different or just have different names (the code indicates that DsFmedian from Table 1 is DsFhmedian in Table 2). However, a subset of filter names is common to both tables. Comparing this common part of Tables 1 and 2 as well as Fig. 12 one comes to the following conclusions. Increasing kurtosis corresponds to better visual quality of edges in filtered images, the best being DsFlsmv and DsFsrad, and the worst DsFwiener2 and DsFlsmisc. In this range there is a good agreement between calculated kurtosis, visual evaluation, and medical expert's satisfaction. The range of higher values of kurtosis cannot be compared with available medical expert's opinion; however, inspection of Fig. 12 indicates further visual improvement of filtered images, which is desirable in case of further processing of images, when certain tissues are to be segmented automatically.

RESULTS

EXPERIMENTS WITH COMPUTER-GENERATED IMAGE

The first example of the calculation of kurtosis relates to the computer-generated image of an artery depicted in Fig. 1. The obvious advantage of this image is its utmost simplicity. In order to get a clear picture of how quality of edges is related to kurtosis, a certain number of frames are selected from the image such that they cover the interesting area in the image. The results given below refer to the average kurtosis of

these frames so the influence of individual frames is reduced. In this way a local kurtosis rather than the global one for the whole image is calculated. The particular arrangement of frames covering one edge in the image is shown in Fig. 1(c). The image of Fig. 1(b), and others to be presented below, were filtered using 14 methods collected and implemented by Loizou as well as the DsNLM filter, and the results for Fig. 1(b) are depicted in Fig. 2.

Table 1. *Ranking of filters for an exemplary image of the carotid artery. The results are ordered according to decreasing kurtosis.*

	Rank	Relative Kurtosis
Original		1
DsFls	1	4.4144
DsFhomog	2	4.1529
DsNCD	3	4.0832
DsFlecasort	4	4.0829
DsFwaveltc	5	3.1433
DsFca	6	2.7545
DsFsrad	7	2.6541
DsFhomo	8	2.4299
DsFad_ver_1	9	2.4116
DsNLM	10	1.9442
DsFad_ver_2	11	1.9269
DsFlsmv	12	1.7418
DsFmedian	13	1.3890
DsFgf4d	14	1.3553
DsFwiener2	15	1.3407
DsFlsmisc	16	0.9959

Table 2. *Ranking of filters for carotid artery images by a neurovascular specialist. The DsFhmedian filter is the same as DsFmedian in Table 1, and the inferior DsFmedian filter in the current table should be dropped. DsFlndif, DsFkuwahara denote Coherent Nonlinear Anisotropic Diffusion Filter and Kuwahara Filter, respectively.*

	Rank	Score
Original		82
DsFlsmv	1	91
DsFhmedian	2	90
DsFkuwahara	3	83
DsFsrad	4	79
DsFad	5	77
DsFmedian	6	76
DsFlndif	7	75
DsFlsmisc	8	74
DsFgf	9	73
DsFwiener	10	71

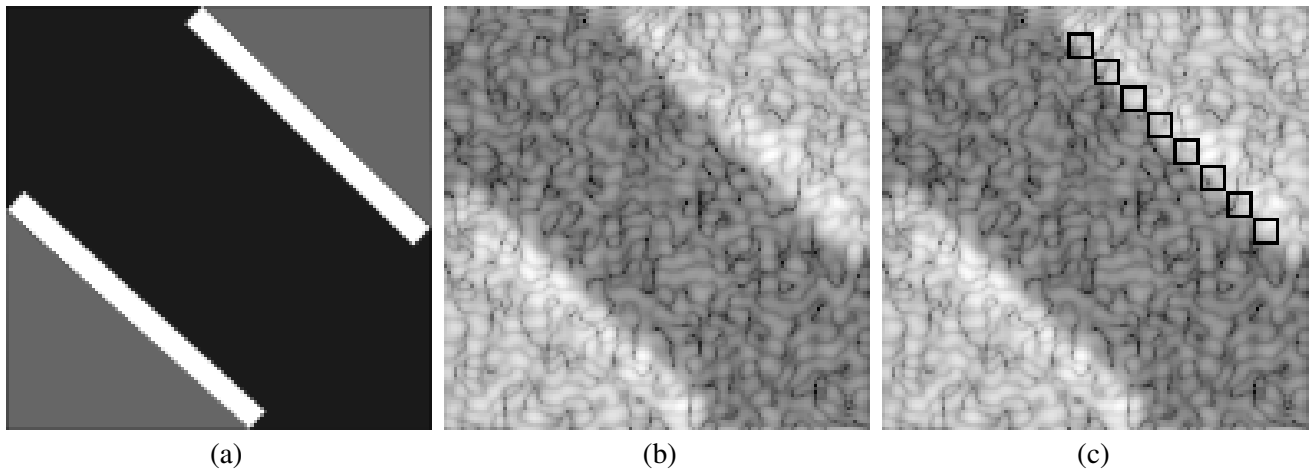


Fig. 1. (a) Original image of the ideal carotid artery without speckles. (b) Computer-generated image with speckles. (c) Frames used for calculating kurtosis.

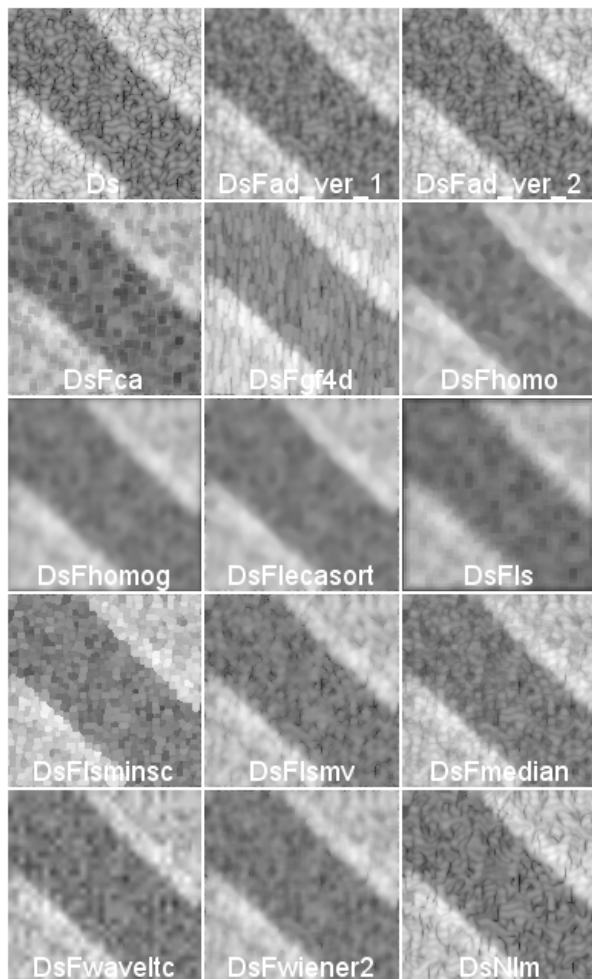


Fig. 2. Results of Loizou and DsNLM filtering for the computer-generated image in Fig. 1(b). The upper left image Ds is a repetition of Fig. 1(b).

The results of calculating kurtosis for the computer-generated image under consideration are shown in column entitled Fig. 1(b) in Table 3.

One can observe in Table 3 that various original images have widely different kurtosis and comparisons may be simplified if we use a relative kurtosis rather than the absolute one. For this purpose all values in any column are divided by kurtosis of the original image in that column, and the results of this operation are presented in Table 4.

Some caution should be exercised when analyzing Tables 3 and 4 since no effort was made to optimize various filtering methods by adjusting their parameters. Whenever possible, simply the default parameters specified in Loizou (2013) were used. In the case of the DsNLM filter the parameters were: search window of side $t = 5$, similarity window of side $f = 2$, and filtering parameter $h = 10$ (Buades *et al.*, 2005; Manjon-Herrera, 2016). For the DsNCD filter the parameters were taken from Nieniewski (2014). Inspection of the column designated Fig. 1(b) in Table 4 reveals that the maximum of relative kurtosis of approx. 1.48 was obtained with the DsFls, and the minimum of 1.09 for DsFlsminsc. Comparing the images one could say the DsFls filter is characterized by relatively sharp edges, whereas for DsFlsminsc the edges are more fuzzy. In summary, the filtering by means of Loizou methods results in a certain increase of kurtosis, but it is hard to actually see better quality when a small change in the value of kurtosis is considered. The human eye is not that sensitive. It may be noteworthy that strictly speaking kurtosis attempts to assess the quality of the edges rather than the overall appearance of the image, and the overall improvement is rather a by-product of improving the edges.

Table 3. Kurtosis for images of Figs. 1(b), 6(a), 9(a), 11(a) with Loizou, DsNLM, and DsNCD filters. The entries for DsFsrad filter were calculated, but the filtered images are visually unsatisfactory and are omitted in Fig. 2 and others.

	Fig. 1(b)	Fig. 6(a)	Fig. 9(a)	Fig. 11(a)	Carotid artery, Ds in Fig. 12
Noisy original	16.3634	25.0361	17.9990	9.9953	21.2495
DsFad_ver_1	19.2793	30.5570	24.2368	14.0957	51.2456
DsFad_ver_2	18.7665	27.2560	21.1573	12.1885	40.9450
DsFca	20.3625	35.7851	28.9888	14.4012	58.5322
DsFgf4d	19.8570	29.6929	21.2105	11.5301	28.7991
DsFhomo	21.0617	35.5243	23.7662	15.4907	51.6340
DsFhomog	22.3698	28.8664	24.6762	21.3712	88.2466
DsFlecasort	21.5025	39.8675	36.0942	18.2440	86.7599
DsFls	24.2665	50.6443	41.7878	21.5752	93.8030
DsFlsminsc	17.8642	20.3367	16.7258	9.5927	21.1627
DsFlsmv	19.0897	32.9382	18.9452	12.6113	37.0118
DsFmedian	18.0387	28.1476	19.1778	11.8536	29.5164
DsFsrad	18.6870	60.7419	18.4322	20.1639	56.3978
DsFwaveltc	18.9927	27.0561	27.4129	18.9772	66.7940
DsFwiener2	19.1694	32.4901	19.6838	12.7861	28.4884
DsNLM	18.2604	40.6894	23.5129	12.4956	41.3123
DsNCD (3 iters.)	22.8319	39.2659	34.3443	18.9755	86.7670

Table 4. Relative kurtosis for images of Figs. 1(b), 6(a), 9(a), 11(a) with Loizou, DsNLM, and DsNCD filters.

	Fig. 1(b)	Fig. 6(a)	Fig. 9(a)	Fig. 11(a)	Carotid artery, Ds in Fig. 12
Noisy original	1	1	1	1	1
DsFad_ver_1	1.1782	1.2205	1.3466	1.4102	2.4116
DsFad_ver_2	1.1469	1.0887	1.1755	1.2194	1.9269
DsFca	1.2444	1.4293	1.6106	1.4408	2.7545
DsFhomo	1.2871	1.4189	1.3204	1.5498	2.4299
DsFhomog	1.3671	1.1530	1.3710	2.1381	4.1529
DsFlecasort	1.3141	1.5924	2.0053	1.8253	4.0829
DsFls	1.4830	2.0229	2.3217	2.1585	4.4144
DsFlsminsc	1.0917	0.8123	0.9293	0.9597	0.9959
DsFlsmv	1.1666	1.3156	1.0526	1.2617	1.7418
DsFmedian	1.1024	1.1243	1.0655	1.1859	1.3890
DsFsrad	1.1420	2.4262	1.0241	2.0173	2.6541
DsFwaveltc	1.1607	1.0807	1.5230	1.8986	3.1433
DsFwiener2	1.1715	1.2977	1.0936	1.2792	1.3407
DsNLM	1.1159	1.6252	1.3063	1.2501	1.9442
DsNCD (3 iters.)	1.3953	1.5684	1.9081	1.8984	4.0832

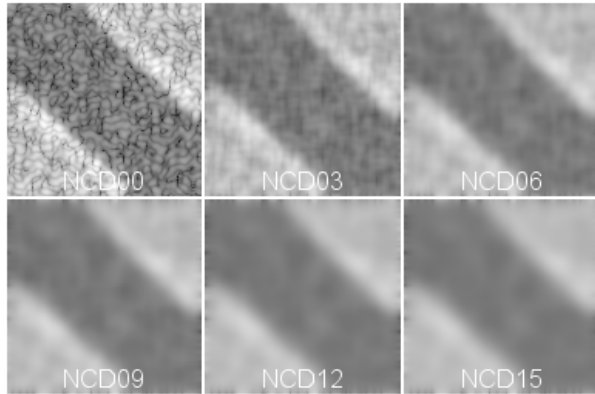


Fig. 3. Results of filtering the computer-generated image from Fig. 1(b) using NCD iterations in the range from 0 through 15 in steps of 3. The upper left image is a repetition from Fig. 1(b).

The aim of Tables 3 and 4 is not to find out which filtering method is the best one, but rather to illustrate what values of kurtosis are obtained and how they compare with the visual evaluation of the respective images. The filtering results can be modified by changing parameters; for example, changing the number of iterations in iterative methods obviously would modify Table 3. Having said this we can make a number of comments on Table 3. The ideal image of Fig. 1(a) has low kurtosis of 9.7167, and kurtosis of the speckled image in Fig. 1(b) is 16.3634 (compare also Fig. 4(a) below). This means that kurtosis increases when speckles are added to a noiseless image. Inspection of Table 4 reveals that in practically all cases kurtosis increases as a result of filtering. The only exception is the DsFlsminsc filter. The Speckle Reducing Anisotropic Diffusion DsFsrad filter from Loizou (2013) is included in Tables 3 and 4; however, the respective filtered images are omitted. The reason for this omission is that Figs. 6(a), 9(a), 11(a) were obtained in an US machine and have a nonuniform background. The DsFsrad filter gave good results for kurtosis of the filtered images; however, these images have too high nonuniformity of background in comparison with originals and cannot be accepted. The case of the computer-generated image does not present this problem.

In another experiment the same image of Fig. 1(b) was filtered by means of the DsNCD filter (Nieniewski and Zajęczkowski, 2014). In this case the conditions were more uniform in the sense that only one filter was used. The filtering results are shown in Fig. 3. The dependence of kurtosis on the number of iterations is illustrated in Fig. 4(a), which shows that kurtosis increases almost linearly with a number of NCD iterations in a relatively wide range.

In the next experiment the computer-generated image of Fig. 1(b) after NCD despeckling was subjected to further filtering by means of the DsFBD filter. The parameters of this filter were taken from Nieniewski and Zajęczkowski (2016). In particular, Fig. 5 illustrates results of three NCD iterations followed by a number of FBD iterations in the range from 0 to 10. It can be observed in Fig. 5 that the DsFBD filter really emphasizes the edges, although the number of iterations should not be too large. Fig. 4(b) shows that the DsFBD filter reduces kurtosis for higher number of FBD iterations, but this is the situation when the speckles have been filtered out.

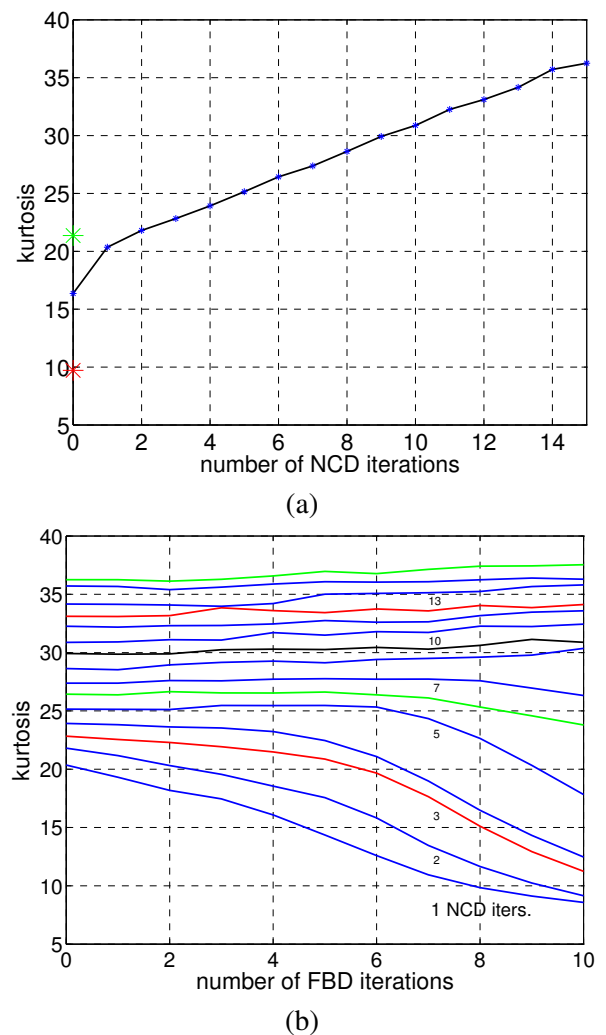


Fig. 4. (a) Kurtosis as a function of number of NCD iterations for image in Fig. 1(b). For comparison, the lower red point on the vertical axis represents kurtosis of image in Fig. 1(a), and the upper green point corresponds to kurtosis for image in Fig. 1(b) filtered with a simple median filter of size 7×7 . (b) Kurtosis as a function of number of FBD iterations for image in Fig. 1(b). The number of prior NCD iterations changes from 1 to 15 and is specified as a parameter for each curve.

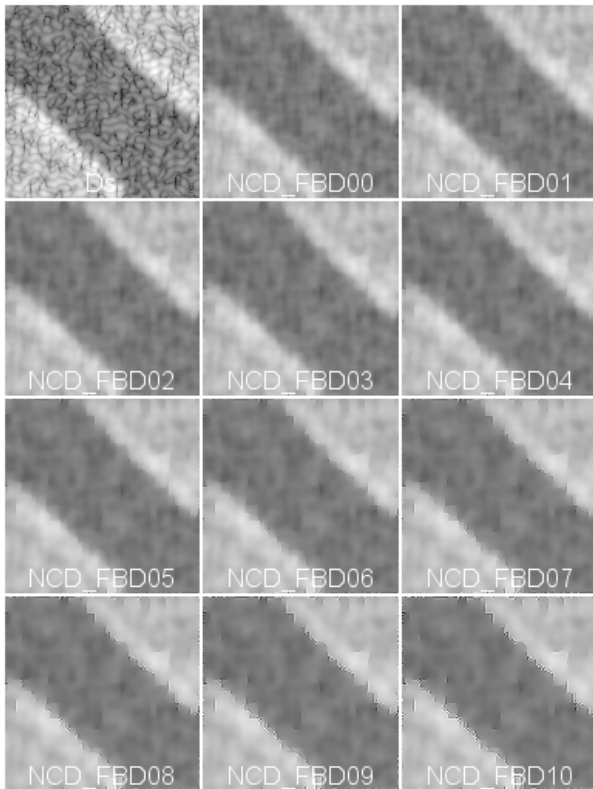


Fig. 5. Results of filtering the computer-generated image from Fig. 1(b) using 3 NCD iterations and subsequent FBD iterations in the range from 0 through 10. The upper left image is a repetition from Fig. 1(b) and the next image is the result of 3 NCD iterations.

EXPERIMENTS WITH IMAGES OF PHANTOMS

The first example is an image of a general purpose phantom shown in Fig. 6(a) in which one of the cylinders imitating a blood vessel is clearly visible and other cylinders on the same level are much less distinct. Fig. 6(b) depicts the zoom of the cylinder area with the assumed frames imposed. The results of Loizou and DsNLM filtering for this case are shown in Fig. 7, and kurtosis values are specified in Tables 3 and 4 in the column entitled Fig. 6(a). Fig. 8 shows kurtosis as a function of a number of NCD iterations for the image in Fig. 6(a). This figure might be compared with kurtosis in Fig. 4. The linearity of the curve is not as good as in Fig. 4(a); nevertheless, the trend obviously is the same.

The second example is the image of the phantom with an imitation of a cyst (there are no speckles inside the cyst) shown in Fig. 9(a). For this image the selected frames are depicted in Fig. 9(b). Results of Loizou and DsNLM filtering are illustrated by Fig. 10(a), and kurtosis values are given in Tables 3 and 4 in the column designated Fig. 9(a).

EXPERIMENTS WITH NATURAL US IMAGE

A natural image of a thyroid, in which an artery is visible, is shown in Fig. 11(a). For this image the selected frames are depicted in Fig. 11(b). Results of Loizou and DsNLM filtering are illustrated by Fig. 10(b), and kurtosis values are given in Tables 3 and 4 in the columns designated Fig. 11(a). Another natural image, of a carotid artery, is also shown and discussed in detail below (compare Fig. 12).

DISCUSSION

KURTOSIS VS OTHER IMAGE QUALITY METRICS

Basically the purpose of the current paper is a development of such image quality metric which would be consistent with visual perception and medical expert's satisfaction and would be preferable in comparison with other metrics. For this reason the actual state of affairs in this area will be considered in more detail. Loizou and Pattichis (2015) present a number of metrics for evaluation of filtered image quality. Their book is not an exclusive source of information in this respect, but it is a recent one and is by far the most comprehensive. The metrics considered in Loizou and Pattichis (2015) measure the quality change between the noisy, unprocessed image and the image filtered by any of the above mentioned methods. The specific names of the metrics and their commonly used abbreviations are the following:

- Universal quality index (Q) defined by the equation

$$Q = \frac{\sigma_{gf}}{\sigma_f \sigma_g} \times \frac{2\bar{f}\bar{g}}{(\bar{f})^2 + (\bar{g})^2} \times \frac{2\sigma_f \sigma_g}{\sigma_f^2 + \sigma_g^2}, \quad (13)$$

where \bar{g} and \bar{f} denote the mean of the original and filtered images, respectively; σ_g and σ_f are standard variations of the original and filtered images, and σ_{gf} denotes the covariance between both images. The value of the universal quality index satisfies the condition $-1 \leq Q \leq 1$, and higher values of Q indicate better agreement of compared images.

- Mean-square error (MSE).
- Root mean-square error (RMSE).
- Minkowski metric (Err).
- Geometric average error (GAE).

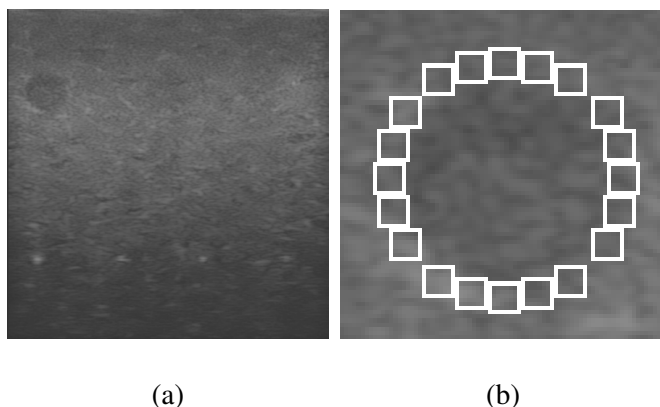


Fig. 6. (a) Phantom image with a cross section of a cylinder in the upper left corner of the image. (b) Frames used for calculating kurtosis.

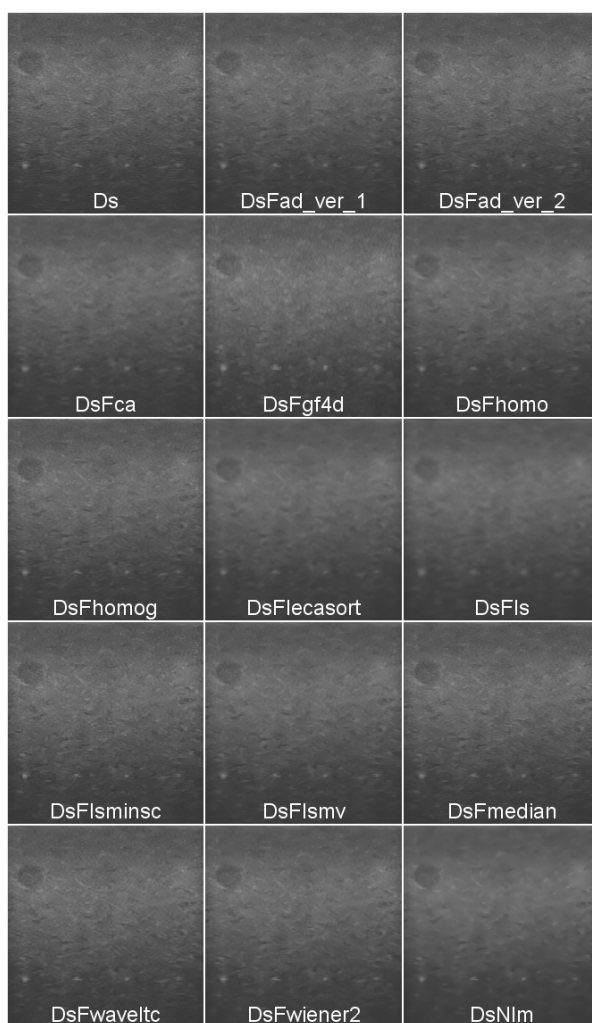


Fig. 7. Results of Loizou and DsNLM filtering the general purpose phantom image in Fig. 6(a).

- Signal-noise-ratio (SNR). is a generalization of the universal quality index Q.
- Peak signal-noise ratio (PSNR).
- Structural similarity index (SSIN or SSIM), which - Speckle index (C) defined by the equation

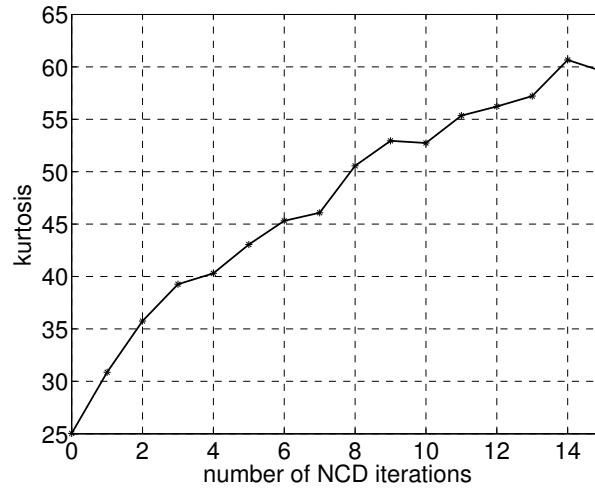


Fig. 8. Kurtosis as a function of number of NCD iterations filtering the general purpose phantom image in Fig. 6(a).

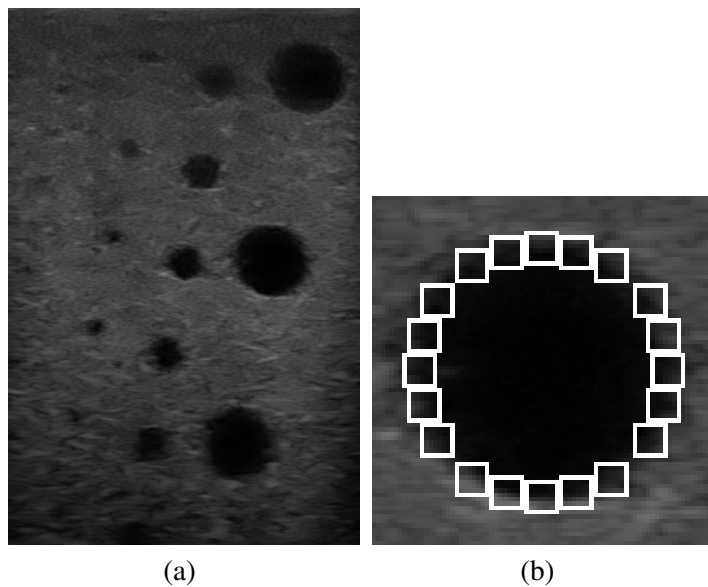


Fig. 9. (a) Phantom image of a cyst to the right of the image center. (b) Frames used for calculating kurtosis.

$$C = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \frac{\sigma_{i,j}^2}{\mu_{i,j}}. \quad (14)$$

The speckle index is an average measure of the amount of speckle present in the image of size $M \times N$, with ratio variance/means computed locally. This is the only metric which uses exclusively the filtered image. The speckle index in general is lower for filtered images in comparison with unfiltered ones.

- Contrast-to-speckle ratio (CSR).
- Average difference between the original and despeckled image (AD).

- Structural content (SC).
- Normalized cross-correlation (NCC).
- Maximum difference (MD).
- Laplacian mean-square error (LMSE).
- Normalized absolute error (NAE).

The details of particular metrics are not of primary importance for us, and besides they can be easily found in the cited literature. The universal quality index is specified more fully here in order to show an example how quality metrics are typically defined. All of the metrics specified in Loizou and Pattichis

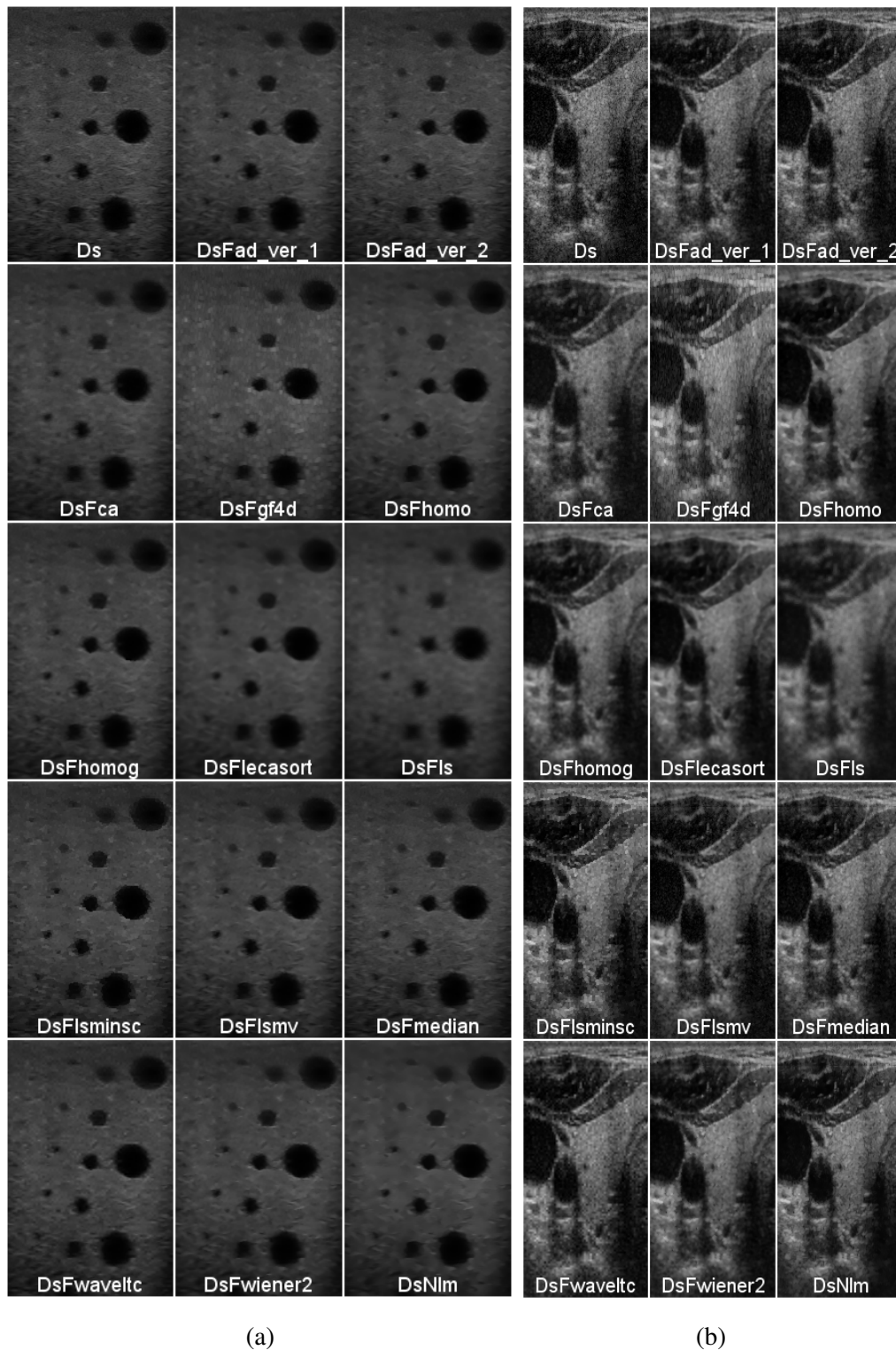


Fig. 10. (a) Results of Loizou and DsNLM filtering the cyst phantom image of Fig. 9(a). (b) Results of Loizou and DsNLM filtering the natural thyroid image of Fig. 11(a).

(2015), with exception of the speckle index, are based on comparison between the reference image and the filtered image. Since the noiseless reference image is unavailable, the original noisy image is taken as a reference. The speckle index, also presented in

more detail above, certainly does not take directly into account all details of the filtered images, such as tissue edges, etc. With lack of ideal noiseless image it is even hard to say in general whether a given metric should increase or decrease in order to get

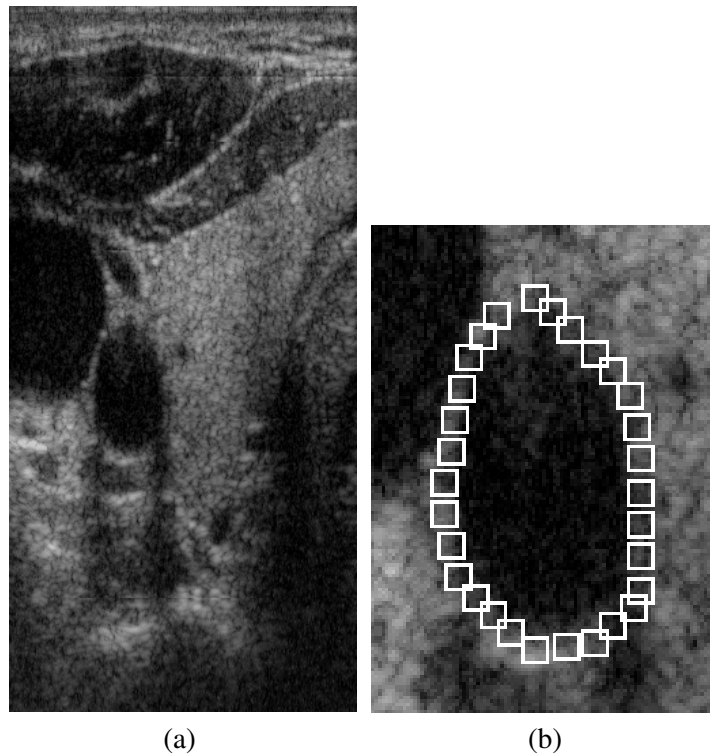


Fig. 11. (a) *Natural image of a thyroid.* (b) *Frames used for calculating kurtosis.*

a better image. Loizou and Pattichis (2015), vol. I, pp. 38 explicitly say that “the quality measures ... do not necessarily correspond to all aspects of the expert’s visual perception ... , but if they are all combined together, and with the subjective tests, may offer a more accurate evaluation result. It is noted that subjective tests are tedious, time consuming, and expensive, and the results depend on the expert’s background, motivation, and other factors.” All the proposed measures cover the visual quality only in part, and it would be desirable to have a measure that is consistent with visual perception as well as medical expert’s satisfaction. The combination of such a large number of metrics is hard to implement since all of them should be included with some weight, and all the weights should be adjusted in such a way that the expert’s satisfaction could be predicted. Of course, one can analyze one metric at a time for a particular example of image of, say, carotid artery or any other anatomical structure, but this is far from universal approach.

In order to be more specific we will discuss some details from Loizou and Pattichis (2015). Three major sets of filtered images are presented in vol. I, section 7.3, of this book. In particular, Figs. 7.4, 7.5, and 7.6 show, respectively, carotid artery images, cardiac images, and carotid artery video. The calculated metrics are contained in Tables 7.3–7.5. An investigative look at this plethora of data confirms

that it may be very hard to combine these metrics in a single measure which would be consistent with medical expert’s opinion. And obviously Loizou and Pattichis (2015) do not suggest any particular solution.

As indicated somewhat loosely in Loizou and Pattichis (2015), vol. I, page 114, and vol. II, page 66, the best visual results as assessed by two experts were obtained for the filters $DsFlsmv$, $DsFlsminsc$ and $DsFkuwahara$. The filters $DsFgf4d$, $DsFad$, $DsFncdif$, and $DsFnldif$ showed good visual results but smoothed the image, losing subtle details and affecting the edges. Table 3.5A in vol. II is more specific and shows the results of the visual evaluation of the original and despeckled images made by a cardiovascular surgeon and a neurovascular specialist. They evaluated 100 ultrasound images of carotid artery before and after filtering: 50 asymptomatic and 50 symptomatic. For the cardiovascular surgeon, the best despeckle filter was the $DsFlsmv$ followed by $DsFgf4d$, $DsFhmedian$, $DsFhomog$ and original. For the neurovascular specialist, the best filter was the $DsFgf4d$, followed by $DsFlsmv$, original, $DsFlsminsc$, and $DsFhmedian$. The difference in the score between the two vascular specialists stems from the fact that the cardiovascular surgeon was primarily interested in the plaque composition and texture evaluation, whereas the neurovascular specialist was interested in the degree of stenosis and the lumen diameter in order to identify the plaque contour.

Table 5. *Ranking of filters by relative kurtosis vs neurovascular specialist. The table is valid for carotid artery images. The results in the left column are ordered according to decreasing kurtosis.*

Relative Kurtosis	Neurovascular specialist
DsFsrاد	DsFlsmv
DsFاد	DsFھmedian
DsFlsmv	DsFsrاد
DsFmedian	DsFاد
DsWiener2	DsFlsminsc
DsFlsminsc	DsWiener

Although the above description seems clear, there is a technical difficulty arising from the fact that the stack of available filters is changing in time. In this situation it turned out more practical to use evaluated common filters taken from Loizou *et al.* (2014) rather than the book of the same author. Based on the Loizou *et al.* (2014) we juxtaposed their results with our own in Table 5. The filters in this table are ranked by kurtosis in the left column, and by the neurovascular specialist in the right column. The positions of the “namesake” filters are relatively close. In fact Table 5 contains the information extracted from Tables 1 and 2, and its contents could be compared with a collection of filtered images in Fig. 12, in which Ds denotes the unfiltered image.

GENERAL EXAMINATION OF RELATIVE KURTOSIS RESULTS

Several noisy US images have been presented. They include computer-generated image, phantom images, and natural images. All of these images have been subjected to speckle filtering using 16 methods. For each filtered image kurtosis has been calculated. Because of different image contents their kurtosis may be quite different. To make comparisons easier it is better to use relative kurtosis, calculated with reference to the original unfiltered image. The relative kurtosis quantitatively describes the effectiveness of the filtering method for a given image. The experiments have shown that with the assumed parameters of the filters the strongest filtering results are obtained for the Linear Scaling Filter DsFls independently of the choice of the original image (compare Table 4). Such filtering may be useful when segmentation and finding the tissue borders are desired, that is further image processing operations are to be performed. From the medical point of view relative kurtosis values in the range up to approx. 1.75 are interesting. It is noteworthy that according to the rightmost column of Table 2 only the top three filters

achieved better medical score than 82 that is value obtained for the original unfiltered image. Based on the above assumptions we collected several filtered images of the carotid artery in Fig. 12 taking into account particularly those for which there are entries both in Tables 1 and in 2. It should be pointed out that we tried to use default filtering parameters; however, changing their values, for example the number of iterations, whenever applicable, one might influence to some degree the ranking of a given filter.

It is obvious from Fig. 12 that there is some oversmoothing of images for highest values of kurtosis, and the best, from medical point of view, images are for medium kurtosis, where the edges are to some degree more pronounced and simultaneously the details of the image are maintained.

ADDITIONAL REMARKS

It is noteworthy that vol. II of Loizou and Pattichis (2015) deals with an even higher number of 65 textural features. However, these features also serve another purpose as they can differentiate between symptomatic and asymptomatic images of carotid artery and are calculated for appropriate image fragments. In this case the quality metric based on multivariate kurtosis can be of some use, but it would be a matter for further investigation.

Although the conducted experiments were conducted with log-compressed images, it is obvious that the same approach could be used for comparisons of various filters dealing with uncompressed images. The comparison of filters for compressed and uncompressed images would necessitate inclusion of the logarithmic operation into the processing channel of relevant filters. This is not a particularly hard problem; however, there are no medical scores available for such a case and it would be necessary to develop a new dataset. Similarly, new varieties of filters keep showing up in the literature all the time and further collection of data might be necessary.

Apart from image quality assessment methods mentioned above there exist some methods which allow one to evaluate certain quality features of images based on one-dimensional kurtosis (Caviedes and Oberti, 2004; Ferzli *et al.*, 2010; Zhang *et al.*, 2011). However, these papers do not consider despeckling evaluation methods nor use multivariate kurtosis.

CONCLUSIONS AND FUTURE DIRECTIONS

The objective of this paper has been to prove the usefulness of relative multivariable kurtosis for

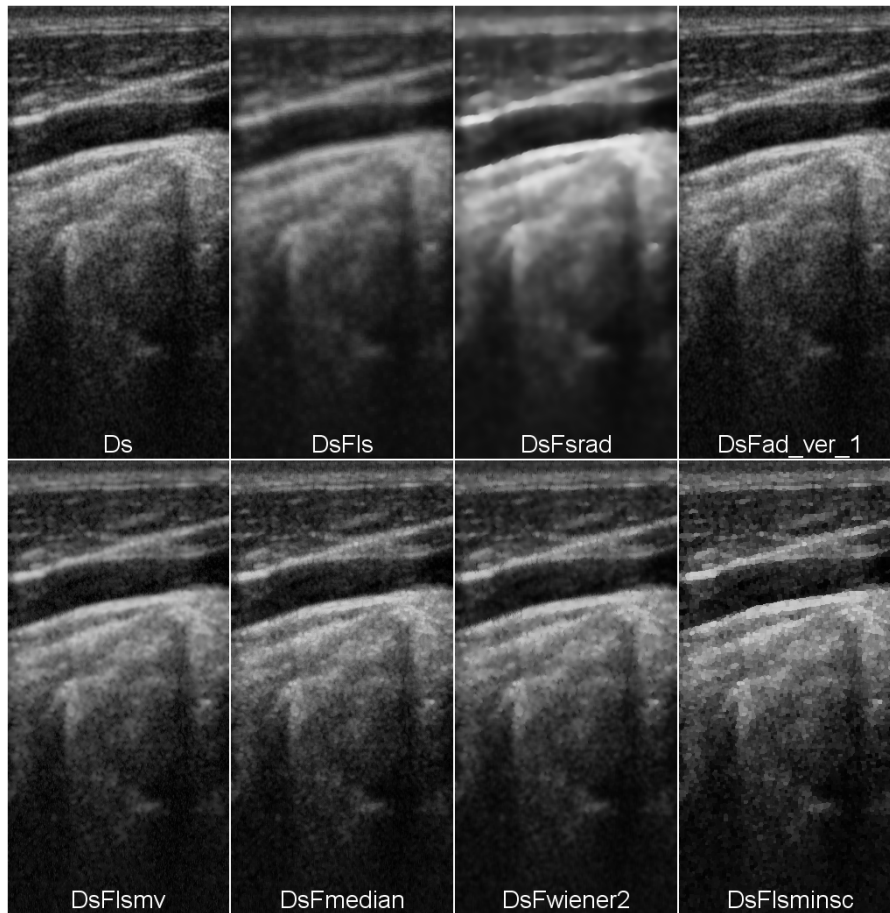


Fig. 12. *Comparison of the filtering results for the carotid artery arranged according to the decreasing kurtosis.*

comparison of US image despeckling methods. The main idea of the presented concept is to directly evaluate the most interesting edges in the image by means of relative multivariable kurtosis. Human vision system is quite sensitive to edges and easily differentiates between sharp and indistinct, imprecise, fuzzy edges. Multivariable kurtosis offers a numerical measure which is consistent with human perception of edges. In a typical situation kurtosis increases due to speckle-removing filtering. Because of the predominant meaning of edges in an image the usefulness of multivariable kurtosis carries over from visual assessment to medical expert's satisfaction with images. In this way we can predict how he/she would evaluate US image quality after filtering. This is particularly important when various filtering methods have to be compared. Multivariable kurtosis does not need any auxiliary images for calculations. The intuitive meaning of kurtosis is superior to numerous image quality metrics, which are much less perception-related and involve the use of a reference image, which by necessity is the unfiltered noisy image. This usage of unfiltered image obscures the meaning of the image

quality metrics described in the literature.

Kurtosis is a highly nonlinear function whose behavior can best be observed by conducting numerical experiments rather than theoretical considerations. A number of despeckling filters have been tested with a computer-generated image, images of phantoms, and natural US images. In addition, it has been shown that multiiterative filtering, such as DsNCD, results in an approximately proportional increase of kurtosis. Once the speckles have been completely removed, one can sharpen the edges by “a touch” of the DsFBD filtering. In this case the tissue's edges become steeper and kurtosis decreases.

In order to have more pronounced numerical results it is recommended to calculate the average of local kurtosis values over a number of frames covering the most interesting edges in the image. It has been found that selecting frames of size 8×8 pixels and transforming these frames into their Fourier counterparts gives meaningful results that can be used for calculating kurtosis and assessing the despeckling effects.

It should be emphasized that the use of the multivariate kurtosis is not restricted to medical images. In fact, multivariate kurtosis can be used for any US images as well as other images contaminated with speckle noise, for example, radar images.

The research reported in this paper can be continued in several directions. The most obvious need is the creation of comprehensive datasets of US images annotated by medical experts as well as creation of an expandable set of filters to which new filters could be added and evaluated as they appear in such a way that their evaluation would not require starting from scratch every time. The more basic problem, however, is a search for alternative evaluation metrics that would not require reference image. Another possible research direction is selection of filters that are optimal from the point of view of anatomical tissues or organs under investigation, from the point of view of a medical specialist, or from the point of view of further automatic image processing.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Prof. Andrzej Nowicki, Institute of Fundamental Technological Research, Warsaw for making available the necessary equipment and helping with acquisition of images for this research.

REFERENCES

- Buades A, Coll B, Morel JM (2005). A non-local algorithm for image denoising. In: Proc IEEE Conf Comput Vis Pattern Recog 2:60-5.
- Caviedes J, Oberti F (2004). A new sharpness metric based on local kurtosis, edge and energy information. *Signal Process: Image Comm* 19:147-61.
- Chandler DM (2013). Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Process* 2013:905685.
- Dansk (2016a). Dansk Fantom Service, <http://www.fantom.dk/1525.html> (Last checked Aug 2016).
- Dansk (2016b). Dansk Fantom Service, <http://www.fantom.dk/571.htm> (Last checked Aug 2016).
- DeCarlo LT (1997). On the meaning and use of kurtosis. *Psych Meth* 2:292-307.
- Ferzli R, Karam LJ (2009). A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE T Image Process* 18:717-28.
- Ferzli R, Girjia L, Ali WS (2010). Efficient implementation of kurtosis based no reference image sharpness metric. In: Proc SPIE 7532: Image Proc Algo Syst VIII 7532:75320E
- Li L, Lin W, Wang X, Yang G, Bahrami K, Kot AC (2016). No-reference image blur assessment based on discrete orthogonal moments. *IEEE T Cybernetics* 46:39-50.
- Lin W, Kuo CC (2011). Perceptual visual quality metrics: a survey. *J Vis Comm Image Represent* 22:297-312.
- Loizou CP, Pattichis CS, Christodoulou CI, Istepanian RS, Pantziaris M, Nicolaidis A (2005). Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. *IEEE T Ultrason Ferr* 52:1653-69.
- Loizou CP, Pattichis CS, Pantziaris M, Tyllis T, Nicolaidis A (2006). Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering. *Med Biol Eng Comput* 44:414-26.
- Loizou CP, Kasparis T, Christodoulides P, Theofanus C, Pantziaris M, Kyriakou E, Pattichis CS (2012). Despeckle filtering in ultrasound video of the common carotid artery. In: Proc 12th IEEE Int Conf Bioinform Bioeng 721-6.
- Loizou CP (2013). Despeckle Filtering Toolbox, http://www.medinfo.cs.ucy.ac.cy/index.php/downloads/toolboxes/10-matlab-software/matlabsoftware_2013.zip (Last checked Aug 2016).
- Loizou CP, Theofanus C, Pantziaris M, Kasparis T (2014). Despeckle filtering software toolbox for ultrasound imaging of the common carotid artery. *Comp Methods Programs Biomed* 114:109-24.
- Loizou CP, Pattichis CS (2015). Despeckle Filtering for Ultrasound Imaging and Video. Vol I: Algorithms and Software, Vol. II: Selected Applications. 2nd Ed. Morgan & Claypool.
- Manjon-Herrera JV (2016). Non-Local Means Filter, <http://www.mathworks.com/matlabcentral/fileexchange/13176-non-local-means-filter> (Last checked Aug 2016).
- Mardia KV (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519-30.
- Mateo JL, Fernández-Caballero A (2009). Finding out general tendencies in speckle noise reduction in ultrasound images. *Expert Syst Appl* 36:7786-97.
- Mittal A, Soundararajan R, Bovik AC (2013). Making a “completely blind” image quality analyzer. *IEEE Signal Process Lett* 20:209-12.
- Narvekar ND, Karam LJ (2011). A no-reference blur metric based on the cumulative probability of blur detection (CPBD). *IEEE T Image Process* 20:2678-83.
- Nieniewski M (2014). Enhancement of despeckled ultrasound images by forward-backward diffusion. In: Proc Int Conf Comput Vis Graphics. Lect Not Comput Sci 8671:454-61.
- Nieniewski M, Zajaczkowski P (2014). Real-time speckle reduction in ultrasound images by means of nonlinear

- coherent diffusion using GPU. In: Proc Int Conf Comput Vis Graphics. Lect Not Comput Sci 8671:462-9.
- Nieniewski M, Zajęczkowski P (2016). Real-time US image enhancement by forward-backward diffusion Using GPU. In: Choraś R, ed. *Image Processing and Communications Challenges 7*. Cham: Springer. 389:177-86.
- Pratt WK (2007). *Digital Image Processing: PIKS Scientific Inside*. 4th Ed. Hoboken, NJ: Wiley.
- Romeu JL, Ozturk A (1993). A comparative study of goodness-of-fit tests for multivariate normality. *J Multivariate Anal* 46:309-34.
- Rosa R, Monteiro FC (2014). Speckle ultrasound image filtering: performance analysis and comparison. In: Proc Computat Vision Medical Image Process IV:65-69, Taylor & Francis.
- Sanches JM, Laine AF, Suri JS (eds.) (2012). *Ultrasound Images. Advances and Applications*. New York: Springer.
- Schäberle W (2005). *Ultrasonography in Vascular Diagnosis*. Berlin: Springer.
- Sheet D (2016). Pseudo B-mode Ultrasound Image Simulator, <http://www.mathworks.com/matlabcentral/fileexchange/34199-pseudo-b-mode-ultrasound-image-simulator> (Last checked Aug 2016).
- Thangavel K, Manavalan R, Aroquiaraj IL (2009). Removal of speckle noise from ultrasound medical image based on special filters: comparative study. *ICGST-GVIP J* 9(3):25-32.
- Thévenaz P (2016). Point Picker, <http://bigwww.epfl.ch/thevenaz/pointpicker/> (Last checked Aug 2016).
- Timm WN (2002). *Applied Multivariate Analysis*. New York: Springer.
- Tong H, Li M, Zhang HJ, Zhang C, He J, Ma WY (2005). Learning no-reference quality metric by examples. In: Proc 11th Int Conf Multimedia Model 247-54.
- Virtanen T, Nuutinen M, Vaahteranoksa M, Oittinen P, Häkkinen J (2015). CID2013: a database for evaluating no-reference image quality assessment algorithms. *IEEE T Image Process* 24:390-402.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004). Image quality assessment: from error visibility to structural similarity algorithms. *IEEE T Image Process* 13:600-12.
- Wang Z, Bovik AC (2006). *Modern Image Quality Assessment*. San Rafael: Morgan & Claypool.
- Wu Q, Li H, Meng F, Ngan KN, Zhu S (2015). No reference image quality assessment metric via multi-domain structural information and piecewise regression. *J Vis Commun Image R* 32:205-16.
- Zhang NF, Postek MT, Larrabee RD, Vlášar AE, Kerry WJ, Jones SN (1999). Image sharpness measurement in the scanning electron microscope – Part III. *Scanning* 21:246-52.
- Zhang J, Ong SH, Le TM (2011). Kurtosis-based no-reference quality assessment of JPEG 2000 images. *Sign Process: Image Comm* 26:13-23.
- Zhang X, Feng X, Wang W, Xue W (2013). Edge strength similarity for image quality assessment. *IEEE Signal Process Lett* 20:319-22.
- Zhu T, Karam L (2014). A no-reference objective image quality metric based on perceptually weighted local noise. *EURASIP J Image Vide* 2014:5.