

Histogram Matching for Visual Ultrasound Image Comparison

Nick Bottenus^{ID}, *Member, IEEE*, Brett C. Byram^{ID}, *Member, IEEE*, and Dongwoon Hyun^{ID}, *Member, IEEE*

Abstract—The widespread development of new ultrasound image formation techniques has created a need for a standardized methodology for comparing the resulting images. Traditional methods of evaluation use quantitative metrics to assess the imaging performance in specific tasks, such as point resolution or lesion detection. Quantitative evaluation is complicated by unconventional new methods and nonlinear transformations of the dynamic range of data and images. Transformation-independent image metrics have been proposed for quantifying task performance. However, clinical ultrasound still relies heavily on visualization and qualitative assessment by expert observers. We propose the use of histogram matching to better assess differences across image formation methods. We briefly demonstrate the technique using a set of sample beamforming methods and discuss the implications of such image processing. We present variations of histogram matching and provide code to encourage the application of this method within the imaging community.

Index Terms—Dynamic range, histograms, image analysis, image quality, ultrasonic imaging.

I. INTRODUCTION

A CONVENTIONAL raw ultrasound image is formed by a delay-and-sum (DAS) beamforming of array channel data followed by envelope detection of the radio frequency or complex quadrature signals to visualize the amplitude-modulated echo signal [1]. The resulting image values can span several orders of magnitude, so the dynamic range is compressed using a function, such as a logarithm or fractional exponentiation. Image postprocessing is often applied to improve the image further, including smoothing background texture, sharpening edges, and enhancing contrast [2].

Historically, quantitative and visual evaluation of ultrasound images has been performed at different stages of this image

formation process. Quantitative metrics are often computed on the raw images, allowing the beamformer to be analyzed in isolation from any image postprocessing and providing certain statistical interpretations about imaging performance. Similarly, visual evaluation of images often benefits from dynamic range compression and image postprocessing to remove artifacts of the image formation process. The vast majority of ultrasound procedures today are performed by human operators viewing compressed and processed images.

Modern ultrasound imaging research has produced a diverse set of imaging techniques that blur the line between the raw image and the displayed image. The widespread availability of open and programmable imaging systems [3] and the improved computational capabilities have led to the proliferation of alternative image reconstruction algorithms. These algorithms range from using models to denoise and recover traditional DAS data [4]–[9], to weighting the DAS image with a correction factor [10]–[14], to replacing the coherent beamsum altogether with a nonlinear image formation method [15]–[20]. The output images for different image formation methods do not necessarily have the same units as DAS; for instance, short-lag spatial coherence [17] produces images in units of correlation that are closer in appearance to the displayed (compressed) B-mode image. It is therefore important to standardize evaluation methods for new algorithm development.

Imaging algorithm development relies heavily on quantitative metrics because they allow different methods to be compared directly in specific imaging tasks. The superiority of a new method is often established by demonstrating improved performance across a panel of metrics in simulation, phantom, and *in vivo* studies. When the ground truth is known, metrics can measure the error directly (e.g., blood velocity, tissue stiffness, cardiac ejection fraction, and the size of a structure). There are also numerous tasks in which the ground truth is unknown or does not exist, where the aim is instead simply to minimize or maximize the metric. Examples include the speckle signal-to-noise ratio (sSNR), lesion conspicuity, and imaging resolution. However, these metrics are often heavily influenced by the types of transformations used to visualize ultrasound images, making it unclear whether the metrics convey a useful comparison between two imaging methods with different units [21]. A quantitative metric was recently proposed for comparing imaging methods with different units in the task of lesion detectability [22], representing an important step toward achieving fair comparisons of image quality.

Manuscript received September 17, 2020; accepted November 2, 2020. Date of publication November 4, 2020; date of current version April 26, 2021. This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering under Grant R01-EB020040 and Grant R01-EB013661. (*Corresponding author: Nick Bottenus.*)

Nick Bottenus is with the Department of Mechanical Engineering, University of Colorado at Boulder, Boulder, CO 80309 USA (e-mail: nick.bottenus@colorado.edu).

Brett C. Byram is with the Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37235 USA.

Dongwoon Hyun is with the Department of Radiology, Stanford University, Stanford, CA 94304 USA.

Digital Object Identifier 10.1109/TUFFC.2020.3035965

In the absence of additional transform-independent metrics for a wider range of tasks, a standardized methodology for visual evaluation takes on greater importance.

Here, we describe a method for separating the evaluation of displayed dynamic range and information content in an image, enabling normalized comparisons between image formation methods. In Section II, we review some common quantitative tools for image analysis and motivate the need for visual assessments. In Section III, we present histogram matching as a tool for fair image comparison and describe several realizations useful for different tasks. In Section IV, we demonstrate the application of this tool to compare several example imaging methods that could otherwise be subject to evaluation bias. In Section V, we highlight the potential pitfalls of this approach and discuss its impact on qualitative assessments.

II. BACKGROUND

A. Common Quantitative Metrics

We begin with a survey of common quantitative metrics. Some metrics simply describe image features, while others have deeper statistical interpretations.

1) *Texture Smoothness*: The characteristics of an image texture are assessed using the sSNR that is defined as

$$\text{sSNR} = \frac{\mu}{\sigma} \quad (1)$$

where μ and σ are the mean and standard deviation of brightness values within a region. When computed on uncompressed B-mode images from diffusely scattering targets (i.e., fully developed speckle), the image values are Rayleigh-distributed and $\text{sSNR} \approx 1.9$ [23]. Higher values of sSNR indicate smoother speckle texture.

2) *Lesion Contrast*: The contrast between two image regions is quantified using the contrast difference (CD) or contrast ratio (CR)

$$\text{CD} = \mu_1 - \mu_2 \quad (2)$$

$$\text{CR} = \frac{\mu_1}{\mu_2} \quad (3)$$

where μ_i is the mean value in region i . CD has units of the underlying image, whereas CR is dimensionless. CD is often used to compare log compressed images that are already in relative units (e.g., decibels), in which case it is identical to the CR of the uncompressed images expressed in the same units.

3) *Lesion Detectability*: The contrast-to-noise ratio (CNR) [24] is measured as

$$\text{CNR} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (4)$$

The CNR describes how visually distinguishable two regions are, combining both contrast and texture variance into one measure. In addition, it has been found that target detectability is related to texture size (i.e., targets are more detectable within finer texture). When computed on raw B-mode images of diffusely scattering targets (i.e., fully developed speckle),

TABLE I
IMAGE METRICS PRESERVED BY TRANSFORMATIONS

Metric	Scale ax	Shift $x + b$	Affine $ax + b$	Monotone $h(x)$
CD	✗	✓	✗	✗
CR	✓	✗	✗	✗
sSNR	✓	✗	✗	✗
FWHM	✓	✗	✗	✗
CNR	✓	✓	✓	✗
gCNR	✓	✓	✓	✓

the CNR, including this resolution term, represents the ideal observer in the task of lesion detection [25], [26].

The generalized CNR (gCNR) between two image regions is computed as the overlap between their histograms

$$\text{gCNR} = 1 - \int_{-\infty}^{\infty} \min\{f(x)g(x)\} dx \quad (5)$$

where f and g are the normalized histograms of two image regions. Unlike CNR, the gCNR provides a nonparametric measure of lesion detectability that is invariant to dynamic range transformations [22]. The gCNR is related to the minimum probability of error by the ideal classification algorithm; higher gCNR indicates a smaller probability of error.

4) *Other Metrics*: Many quantitative metrics exist for other imaging tasks as well. For instance, in a linear imaging system, the full-width at half-maximum (FWHM) of a point target describes the system resolution, whereas the sidelobes of the point spread function describe the level and spatial extent of energy that is expected to produce off-axis scattering. Cystic resolution represents this spatial distribution of the point spread function by predicting target detectability as a function of lesion radius [27] and signal-to-noise ratio [28]. Contrast linearity [21], [29] has been proposed as a way to describe the accuracy of an imaging method in representing the native target contrast relative to the background irrespective of global transformations.

B. Impact of Dynamic Range Transformations

Quantitative metrics are often computed on the uncompressed B-Mode images, while the displayed images have typically undergone dynamic range compression, a type of monotonic transformation. However, most of the aforementioned metrics are not preserved under general monotonic transformations. Table I shows a list of several of these metrics and their theoretical properties under scaling, shifting, affine, and monotonic operations on an image with pixel values x . Among these, only the gCNR is preserved under all monotonic transformations.

Furthermore, the ranking of imaging methods from best to worst based on these metrics is generally not preserved under monotonic transformations. An imaging method that is quantitatively superior to raw images may be worse when the images have been compressed. Consequently, there may be a mismatch in quantitative metrics computed on raw images versus the actual displayed images. Fig. 1 shows a simple example of how the CD values and rankings are affected by a simple dynamic range transformation. Note that this effect is exacerbated when image formation or postprocessing applies a different effective transformation to each image. For

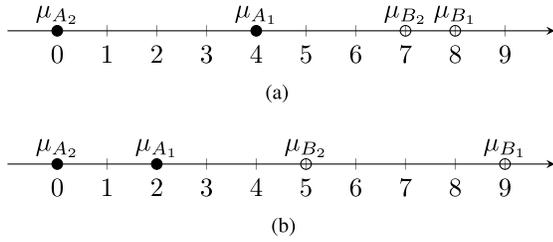


Fig. 1. Simple example of how a monotonic transformation can affect contrast. Let μ_{A_1} and μ_{A_2} denote the mean values of image A in regions 1 and 2, shown plotted on a real number line. The CD of method A is $\mu_{A_1} - \mu_{A_2}$, i.e., the distance between the points. (a) Before monotonic transformation, A has greater contrast than B . (b) After monotonic transformation, B has greater contrast than A . Not only are the contrasts not preserved for A and B , the relative *ranking* of A and B contrast is not preserved.

example, it is difficult to compare a log-compressed DAS image produced in a research lab to a commercial scanner image using an unknown compression and contrast-enhancing grayscale transformation.

C. Normalized Image Quality Evaluation

These observations strongly motivate the need to evaluate images in the domains of their respective target applications. While quantitative metrics on raw images may be appropriate for applications such as computer-aided diagnosis or ideal observer analysis, imaging applications intended for human observers should also be evaluated using the displayed images. Clinical reader assessments are a common tool for evaluating the diagnostic value of new methods, often in the context of a particular clinical task. In cases when extensive reader studies are not feasible, simple quantitative metrics are needed to identify promising methods. However, dynamic range transformations can have a significant influence on quantitative metrics and may break the assumptions made by ideal observers [21], [22]. Whether the visual assessment is performed by clinical readers or quantitative metrics, we require a robust and rigorous methodology for comparing new methods.

We propose the use of histogram matching to overcome the bias introduced by dynamic range transformations. Histogram matching is a family of techniques to ensure that two or more images are properly normalized for presentation [30]. This tool is often used in photography to compensate for the transfer function of an imaging sensor or varying lighting conditions. In the following, we describe the application of this tool to ultrasound images and its impact on common imaging tasks, including quantitative assessment.

III. METHODS

In the following, we present histogram matching for image normalization. In all cases, we consider an input image X and a reference image Y . The goal is to find a transformation of X that best matches Y . As we will show, the choice of transformation has a significant impact on the resulting perception and image quality metrics.

A. Partial Histogram Matching

The simplest transformations for histogram matching are the scaling and shifting operations

$$X_{\text{scale}} = aX \quad (6)$$

$$X_{\text{shift}} = X + b. \quad (7)$$

These operations transform the histogram of X . Scaling affects the dynamic range of the image, whereas shifting the mean affects the overall brightness of the image. When combined, the resulting operation is referred to as an affine transformation, i.e., $X_{\text{affine}} = aX + b$.

Histogram matching with affine transformations seeks a and b that make the transformed image $aX + b$ most similar to Y . In this work, we choose a and b that match the mean μ and variance σ^2 of X to that of Y

$$a = \frac{\sigma_Y}{\sigma_X} \quad (8)$$

$$b = \mu_Y - a\mu_X. \quad (9)$$

This mean and variance matching method is easy to compute because it relies on simple descriptive statistics and has the effect of maximizing the luminance and contrast terms of the structural similarity (SSIM) [31], a popular visual similarity metric for comparing images. For the same reason, it is somewhat robust to outliers in the distribution. Note that this method will only perfectly match two distributions that vary by at most these two moments, such as two images from Rayleigh distributions or Gaussian distributions or two images differing by a linear transformation. An alternative matching scheme is to minimize the L_2 norm

$$\min_{a,b \in \mathbb{R}} \|aX + b - Y\|_2 \quad (10)$$

which as derived in the Appendix and produces the weights

$$a = \frac{\sigma_{XY}^2}{\sigma_X^2} \quad (11)$$

$$b = \mu_Y - a\mu_X \quad (12)$$

where σ_{XY}^2 is the covariance of X and Y . It should be noted that other functions, such as the L_1 norm, can also be used for optimization but may not have a closed-form solution for the affine coefficients.

Partial histogram matching fails to match the overall appearance of two images when their distributions significantly differ in shape (e.g., significant differences in higher order moments such as skew and kurtosis). This also includes mixtures of distributions where the statistics of the overall data may be matched, while the component distributions may not be.

B. Full Histogram Matching

For distributions not fully matched by the partial method, closer histogram matching can be achieved by considering a wider class of image transformations, whereas partial histogram matching was restricted to the class of affine transformations, “full” histogram matching seeks any monotonic transformation that best matches the input probability density function (pdf) to the reference pdf. Unlike in the partial case,

full histogram matching seeks to completely match the target distribution to the reference.

In practice, full histogram matching is accomplished by finding the mapping h from the input cumulative distribution function (CDF) F_X to the reference CDF F_Y [32], [33]

$$h[i] = j \text{ such that } F_X[i] = F_Y[j] \quad (13)$$

for discretely binned data. Interpolation is used between the bins to most closely match the overall distribution. The mapping function is then applied to the target distribution using interpolation to map individual pixel values

$$X_{\text{full}} = h(X). \quad (14)$$

The result should be that the target histogram matches closely with the reference histogram, depending on the resolution of binning.

C. Variations of Histogram Matching

There are important considerations for histogram matching.

1) *Binning*: The limit of the binning process is to use individual pixels (1 pixel/bin). This “point” match can be simplified to rank-ordering the pixels of each image using a direct one-to-one mapping to replace the target pixel values. This process will exactly match the reference distribution but is particularly susceptible to outliers in the reference data since all values are replicated in the target.

2) *Clipping*: Ultrasound data are commonly clipped for display by truncating the lower and/or upper range of the pixel values. For example, in a log-compressed B-mode image, the lower end of the distribution extends as far as the noise floor of the system allows—well beyond useful information content. Clipping values lower than some threshold place the center of the distribution closer to the middle of the displayed dynamic range. In the above mean and variance matching, it is essential to operate on the whole image distribution (without clipping) so that the calculated statistics best represent the underlying data. This is not required for the full histogram matching approach, where the same percentage of the distribution will simply be mapped to the clipped value. For the case of matching a new image to a “black-box” image, with access only to the display data with clipping applied, a full histogram matching approach may therefore be preferable.

3) *ROI-Based Histogram Matching*: If the reference image has extreme values (e.g., bright points or dark lesions), histogram matching would alter the entire target image to compensate. In these cases, it may be useful to determine a transformation based on only a particular region of interest (ROI) and then to apply that transformation to the whole image. For example, a homogeneous speckle ROI within the image is expected to have a well-behaved distribution (log-Rayleigh for a log-compressed B-mode image). Histogram matching based on this ROI should produce a well-matched background “look and feel” to the images. Differences in other structures, such as noise and acoustic clutter within hypoechoic lesions, can then be assessed between the images. Alternatively, images could be matched to an analytical or reference distribution if a homogeneous ROI cannot be identified. When

using an ROI, it may be possible that the full image contains pixel values outside the range of those in the ROI and therefore outside the mapping function. These values can be clipped to the mapping range or we may employ linear extension of the mapping function to try to preserve the full dynamic range of the original target data. For example, a point target may have a brighter value than is present in the ROI speckle background and requires extrapolation to remain brighter in the matched image.

The abovementioned methods describe global dynamic range transformations and are the main focus of this work. Spatially adaptive matches are also possible by calculating the mapping function separately for multiple ROIs [34], but we will consider these to be more as image postprocessing methods since they actually change the spatial information content of the image.

D. Test Cases for Evaluation

Sample data were acquired using the Verasonics Vantage 256 (Verasonics Inc., Kirkland, WA, USA) using the P4-2v phased array transducer. Channel data were stored for individual element transmissions with center frequency 4.5 MHz for each element on the array. Synthetic aperture focusing was applied to coherently sum together the transmissions from all elements at each point in the image [35]. The ATS 549 general and small parts phantom (CIRS Inc., Norfolk, VA, USA) was imaged to visualize speckle background, contrast lesion targets (−15 and −6 dB), and point targets.

To illustrate the impact of histogram matching across a wide spectrum of signal and image processing, we selected four representative imaging methods.

- 1) *Conventional B-Mode*: Image formation by summation of the focused receive channel data, envelope detection (using the Hilbert transform), and log compression. This is the baseline case to compare against.
- 2) *Square Law Detector*: Similar to conventional B-mode but envelope detection is performed by squaring the RF signal and low pass filtering, resulting in values approximately squared compared to the conventional B-mode (or doubled after log compression). This represents a case with little to no change in underlying information and only a simple (linear) dynamic range alteration.
- 3) *Receive Compounded*: Incoherent spatial compounding, used to reduce speckle texture, by applying envelope detection to each receive channel individually before summation and log compression [36]. This represents a case of a true change in underlying information while retaining a similar but reduced dynamic range.
- 4) *Short-Lag Spatial Coherence (SLSC)*: Image formation by averaging of complex cross correlation of receive channel signals up to ten elements apart after synthetic focusing [37], [38]. Unlike the abovementioned methods that are shown on a logarithmic scale, SLSC images are typically displayed on a linear scale between [0, 1]. This represents a case of a true change in underlying information and image display units (i.e., based on correlation instead of signal amplitude).

TABLE II

FWHM RESOLUTION (MILLIMETERS) FOR POINT TARGETS IN FIG. 5

Matching	B-mode	Square	Spat. comp.	SLSC
Original	2.04	1.44	2.72	4.72
Partial	2.04	2.04	2.40	4.22
Partial (ROI)	2.04	2.02	1.88	4.36
Full	2.04	2.04	2.84	1.62
Full (ROI)	2.04	2.24	2.20	3.62

Histogram matching was performed using 256 bins in all cases. The partial histogram matching method was implemented using the mean and variance matching of (8) and (9).

In order to make the histogram matching methods used in this work accessible to others in the imaging community, a MATLAB implementation and sample phantom data have been made available at https://github.com/nbottenus/histogram_matching (DOI: 10.5281/zenodo.4124190).

IV. RESULTS

As a tool for fair qualitative image comparison, histogram matching is perhaps best evaluated by showing the images themselves for the various processed data sets. Fig. 2 shows the performance and characteristics of the partial and full histogram matching techniques both when applied based on the entire image and based on only a speckle ROI in the middle of the image. All matched images are shown on the same dynamic range. For the cases that determined a histogram transfer function based on the ROI, any values outside the range of the transfer function were calculated using linear extrapolation.

The original images show very different image characteristics. Compared with the conventional B-mode image, the square law image shows a darker background texture and finer point targets that stand out from the background. This appearance would fool both qualitative and quantitative assessments to say that resolution and contrast have improved. Spatial compounding shows the opposite effect, reducing texture especially within the darker lesion but with a haze over the image and reduced prominence of the point targets. SLSC, displayed on a linear scale rather than logarithmic, shows a bright background with high texture variance but extreme contrast of the lesions. SLSC also uniquely shows a small defect in the imaging phantom around 30 mm depth as a dark band, representing low spatial coherence.

The histograms for these images are shown in Fig. 3. The bottom of Fig. 3 shows the raw data as it was clipped and normalized for display, truncating the square law image's distribution and showing how the logarithmic images compare to the linear SLSC image. The clipping of the square law image produces the appearance of the point targets against a black background rather than showing any of the background texture. The spatially compounded image shows a compressed lower range, raising the image mean and creating the apparent haze. This compression potentially biases an observer to believe that lesion detectability has been reduced despite an accompanying smoothing of the background texture. The SLSC image shows the widest range including clipping on the low end, explaining the extreme appearance of the texture in the image and the dark

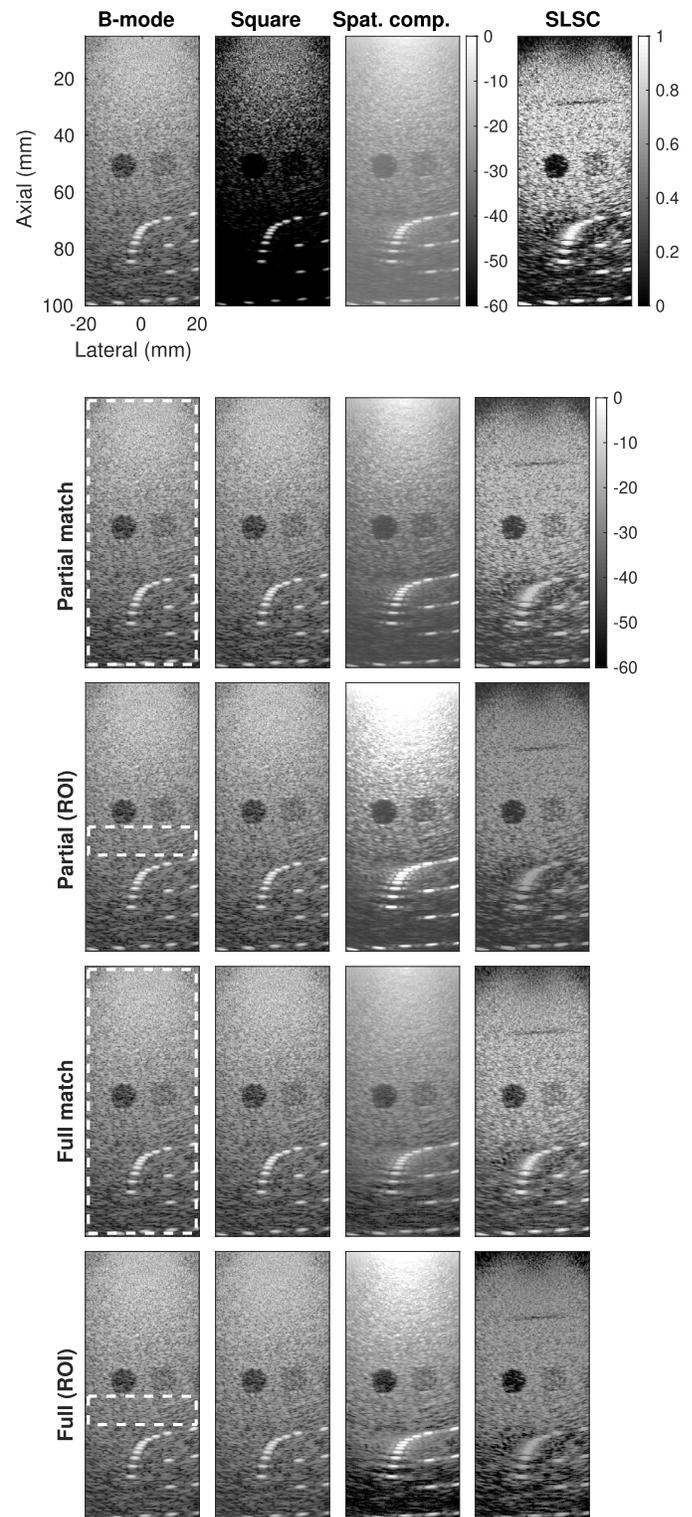


Fig. 2. Top row: original images for four image formation methods—conventional B-mode, square-law detector, receive spatial compounding, and SLSC. SLSC is shown on a linear scale, whereas the others are shown on a logarithmic scale (dB). Bottom rows: images after partial or full histogram matching, using either the full image or a speckle ROI (indicated by white dashed box). All matched images are shown on the same logarithmic scale.

lesion. This stretching of the dynamic range potentially biases visual comparisons against the B-mode image to conclude that the lesion detectability has improved despite an accompanying increase in the background texture variance.

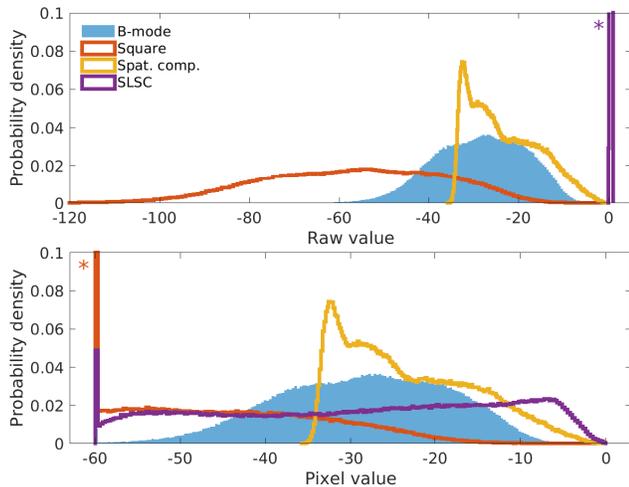


Fig. 3. Histograms for the original images (top) using raw image data values (logarithmic except for the linearly scaled SLSC image) and (bottom) after clipping of data to the displayed -60 -dB dynamic range of the B-mode in Fig. 2. The SLSC image, normally displayed on a linear scale from 0 to 1, has been normalized to the displayed range for comparison. Asterisks indicate that the histogram extends past the displayed Y-axis.

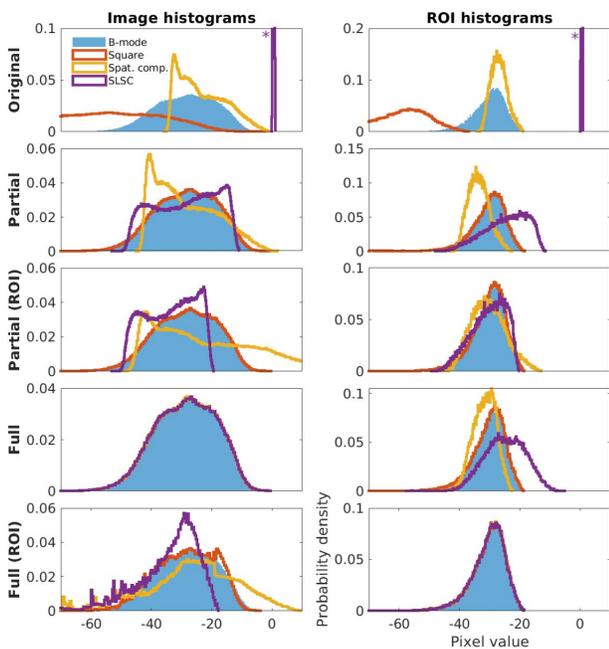


Fig. 4. Results of the proposed histogram matching strategies corresponding to Fig. 2 evaluated for (left) the entire image and (right) the speckle ROI only. Asterisks indicate that the histogram extends past the displayed Y-axis. All histograms are shown with pixel value from $[-70, 10]$, although the histograms may extend further.

Fig. 4 shows the image histograms after the proposed histogram matching strategies were applied. For each, histograms are shown for the entire image and only the homogeneous speckle ROI identified in the images. In the original B-mode image, the background speckle region appears as a well-defined log-Rayleigh distribution as expected. In all matching cases, the square law data are almost exactly matched to the conventional B-mode. After partial matching, all four histograms have much more overlap. However, the differing shapes of the overall distributions for spatial compounding

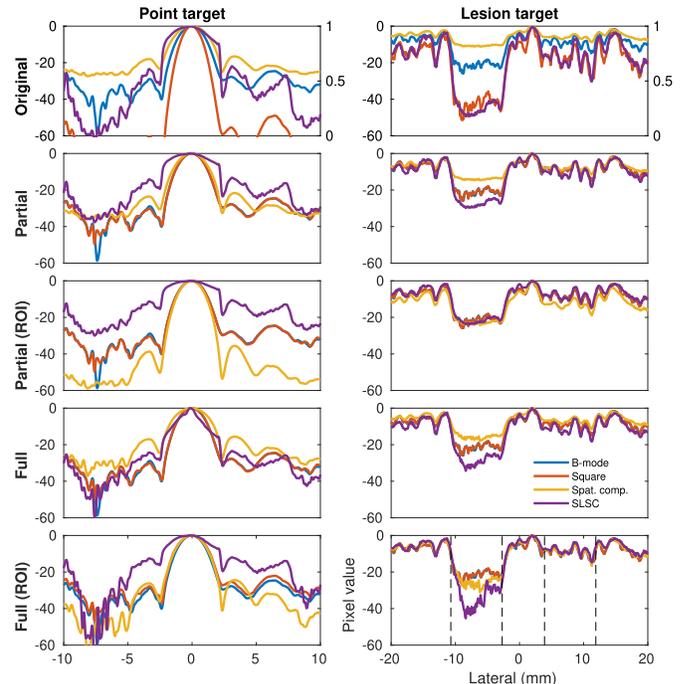


Fig. 5. Four image formation methods compared in each plot for each matching method corresponding to Fig. 2. For the original images, SLSC is shown on a scale from 0 to 1. Left: point target located at $(-1, 84)$ mm, normalized to the maximum brightness. Right: contrast lesion targets located at 50 mm depth, normalized to the maximum brightness. Lesion targets, nominally -15 and -6 dB, are approximately marked by black dashed lines in the bottom-right plot.

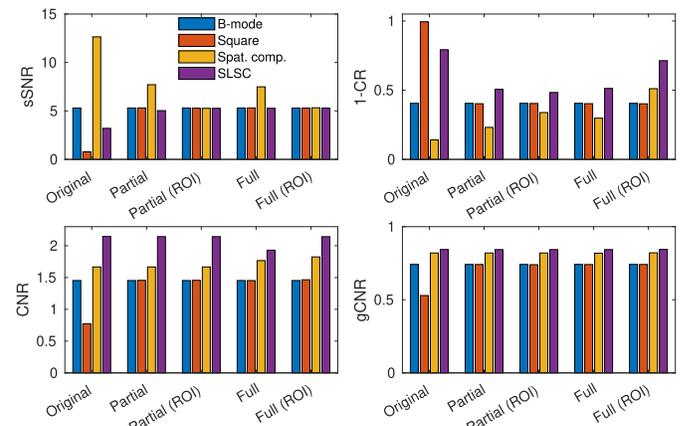


Fig. 6. Image metrics calculated for the four image formation methods across matching methods for a -15 -dB lesion and speckle background. Metrics are calculated using clipped and normalized image data as in the bottom of Fig. 3. Contrast is represented as $1 - CR$ such that a higher bar represents better contrast.

SLSC result in differing means and variances for the speckle background. Conversely, partial matching of the distributions within the speckle ROI leaves large differences in the overall image. The square law data after logarithmic compression are approximately a linear transform of the B-mode data and is therefore matched by the partial (affine) technique for both the entire image and ROI. More complicated transforms, including nonlinear grayscale transforms such as sigmoidal compression for contrast enhancement [22], [39], require full matching. Full histogram matching, as expected, produces near-perfect overlap for the entire image and speckle background when referencing the entire image and ROI, respectively. Matching

the entire image leaves differences in the mean of the speckle background similar to the partial method, but the shapes of the distributions are more similar. Matching the speckle ROI leaves differences in the entire image, most notably in the low end of the SLSC image as it is stretched out to match the reference.

As discussed previously, imaging algorithms are often evaluated using point and lesion targets. These targets are shown in cross section before and after image matching in Fig. 5. The measured FWHM (-6 dB for logarithmic images and 0.5 amplitude for linear images) for each case is given in Table II. Lesion targets were averaged over a 2 -mm axial region to reduce noise. It is evident that without image matching, the interpretation of these two targets is easily distorted by the original image display choices. For example, the square-law detector shows both the best resolution and lesion contrast (at the expense of texture variance, only somewhat evident from these plots due to axial averaging). The reduced dynamic range of the spatial compounding method would make it seem the worst for both targets. However, after applying different matching schemes, the apparent performance of the image formation methods varies drastically. After partial matching based on the speckle ROI, spatial compounding appears to have the best resolution and lowest sidelobes while performing comparably to the other methods in the lesion. It is evident that the center of the point target in SLSC becomes distorted under full histogram matching.

Four common image metrics used to evaluate imaging algorithms—sSNR, CR, CNR, and gCNR—are shown in Fig. 6 under the different matching strategies. Before matching, the speckle texture in the ROI (reflected by sSNR) is the smoothest in the spatially compounded image and most variable in the square law image. Note that partial matching by design matches the sSNR (μ/σ) of the chosen ROI exactly. After matching the whole image, compounding was observed to still produce the smoothest speckle texture, whereas the other images were comparable. Lesion metrics were computed for the -15 -dB lesion relative to the speckle ROI. The contrast in the original square law image was greatly exaggerated due to the stretched dynamic range. After matching, SLSC produced the greatest contrast (expressed as $1 - CR$), while spatial compounding showed varying contrast depending on matching strategy. CNR was invariant to partial matching, as described in Table I with the exception of the square law image due to clipping. The difference in CNR between compounding and SLSC remained even after full matching but was slightly reduced. The gCNR was invariant to image matching, as described in Table I with the exception of the square law image due to clipping.

V. DISCUSSION

Histogram matching provides a systematic approach for placing different imaging methods on a level playing field for visual evaluation. We have presented several variations, providing control over the matching transformation (partial versus full) and the matching region (local ROI versus whole image); one can further vary the matching objective function (mean and variance matching versus L_2) and compensation

for dynamic range clipping. As shown in Fig. 2, histogram matching gives seemingly disparate imaging methods a similar appearance and reverses applied dynamic range transformations.

We expect that the most common use will be full histogram matching to a reference B-mode image using a homogeneous speckle ROI. B-mode speckle represents a known distribution that contains most of the echogenicity values expected in a typical clinical image and should enable a well-behaved transformation function between distributions of differing shapes. Image interpretation is then performed, knowing that the speckle texture has been matched so that differences can be identified in target structures such as lesions. However, it should be recognized that point targets and anechoic structures may fall outside this distribution and are therefore limited by the chosen extrapolation of the histogram matching function. Partial matching may be more appropriate in circumstances where the structure is included in the matching ROI or other situations where the overall shape of the histogram should be preserved. The method of matching should be clearly stated and justified to allow for appropriate interpretation of images. In some cases, it may be desirable to derive a different matching condition than presented here. For example, while the mean and variance affine match preserves sSNR, it is also possible to derive a transform that preserves lesion contrast that would highlight differences in the background texture.

There is not a universal “ideal” distribution for image matching. The structural content and, therefore, histograms of different clinical images vary widely between the targets. For example, the liver appears as a largely homogeneous speckle, whereas the heart contains both brightly scattering tissue and low-scattering blood volume. Clinical imaging systems provide application-specific presets, which employ nonlinear grayscale transfer curves that monotonically transform the displayed dynamic range to highlight structures of interest, maximize contrast between targets, and reject noise. These adjustments, and therefore, the ideal distribution for histogram matching vary widely based on imaging target and task.

The direction of image matching must also be selected with care. In Fig. 2, all images were matched to the compressed B-mode image. This is a natural choice for evaluating novel images in clinical settings where conventional B-mode is familiar. However, a goal of many novel imaging approaches is to detect features that are not visible in the B-mode image; features such as hypoechoic targets obscured by acoustic clutter may not be well-captured by the B-mode dynamic range and consequently may be lost during image matching. In these cases, it may be beneficial to perform matching in the other direction, i.e., the conventional B-mode to the novel image. This is especially important when images are clipped for display and novel information may be lost below a threshold that was selected to minimize noise in the original image. In the case of task-based metrics, improved performance across multiple metrics after matching in both directions could provide compelling evidence of superiority (note that for transform-invariant metrics such as the gCNR, both directions are equivalent).

The histogram bins must be selected so as to allow proper observation of the dynamics of both histograms in order to enable an accurate mapping from one to the other. For instance, raw B-mode images will have a greater concentration of values near zero due to their high dynamic range, in which case the bins should be selected either on a logarithmic scale or after an appropriate dynamic range compression. It is also possible that quantization of the histogram bins may introduce some slight inconsistencies in the image. For the data in this study, we found that 256 bins achieved consistent matching results, although fewer bins may be sufficient.

Evaluation, quantitative or qualitative, is best done on matched images at a stage that isolates the desired comparison. For example, comparison of beamformers as in this article may best be performed on images formed before additional postprocessing. Similarly, image postprocessing should be assessed based on the same input data where possible. The interaction of each beamformer with a given postprocessing method, such as an adaptive speckle reduction or edge enhancement, may vary, compounding the difficulty of interpretation if assessed together (although assessment with postprocessing may ultimately be required for practical application [40]). Special attention should be paid to beamforming with neural networks [20] where a trained network may include some effective image postprocessing not separable from the beamforming step. The histogram matching process is fundamentally the same in all cases, although the ROI to match may change based on the selected evaluation task.

Figs. 5 and 6 confirm the volatility of sSNR, lesion contrast, point target FWHM, and CNR under different dynamic range transformations. In particular, FWHM is easily confounded by image formation and processing. The rescaling of pixel values in the image can create an apparent change in resolution without reflecting the underlying information content of the data. Sparrow's criterion [41], which measures the spacing at which two points form a flat plateau, is an alternative measure that unaffected by such dynamic range transformations. However, it still does not capture the full spatial distribution of energy in the tails of the point spread function and is more difficult to experimentally realize. These sensitivities underscore the critical need for a holistic view of image quality that combines task-based evaluation metrics with visual inspections and expert assessments. Histogram matching is a tool to minimize unintentional biases inherent to different imaging methods and represents one step toward achieving this goal.

VI. CONCLUSION

We have presented a methodology for the fair visual comparison of different ultrasound imaging techniques. Histogram matching provides a systematic approach for minimizing the visual differences between two images according to their histograms and is meant to complement existing quantitative metrics. We have demonstrated partial (affine) and full (monotonic) histogram matching methods based on ROI and whole-image comparisons and showed their effects on visualization as well as on traditional image metrics. Histogram

matching is an integral tool for a comprehensive evaluation of imaging methods.

APPENDIX

A. L2 Affine Matching

The L2 affine histogram match can be posed as minimizing the two-norm of the difference between $a\mathbf{x} + b\mathbf{1}$ and \mathbf{y}

$$J = \|a\mathbf{x} + b\mathbf{1} - \mathbf{y}\|_2^2 \quad (15)$$

$$= (a\mathbf{x} + b\mathbf{1} - \mathbf{y})^T (a\mathbf{x} + b\mathbf{1} - \mathbf{y}) \quad (16)$$

$$= a^2 \mathbf{x}^T \mathbf{x} + b^2 N + \mathbf{y}^T \mathbf{y} + 2abN\mu_{\mathbf{x}} - 2a\mathbf{x}^T \mathbf{y} - 2bN\mu_{\mathbf{y}} \quad (17)$$

noting that $\mathbf{1}^T \mathbf{1} = N$, $\mathbf{x}^T \mathbf{1} = N\mu_{\mathbf{x}}$, and $\mathbf{y}^T \mathbf{1} = N\mu_{\mathbf{y}}$. The partial derivative with respect to a is equal to zero when

$$\frac{\partial J}{\partial a} = 0 = 2a\mathbf{x}^T \mathbf{x} + 2bN\mu_{\mathbf{x}} - 2\mathbf{x}^T \mathbf{y} \quad (18)$$

$$a = \frac{\mathbf{x}^T \mathbf{y} - bN\mu_{\mathbf{x}}}{\mathbf{x}^T \mathbf{x}} \quad (19)$$

and the partial derivative with respect to b is equal to zero when

$$\frac{\partial J}{\partial b} = 0 = 2bN + 2aN\mu_{\mathbf{x}} - 2N\mu_{\mathbf{y}} \quad (20)$$

$$b = \mu_{\mathbf{y}} - a\mu_{\mathbf{x}}. \quad (21)$$

Plug (21) into (19)

$$a = \frac{\mathbf{x}^T \mathbf{y} - (\mu_{\mathbf{y}} - a\mu_{\mathbf{x}})N\mu_{\mathbf{x}}}{\mathbf{x}^T \mathbf{x}} \quad (22)$$

$$a = \frac{\mathbf{x}^T \mathbf{y} - N\mu_{\mathbf{x}}\mu_{\mathbf{y}}}{\mathbf{x}^T \mathbf{x} - N\mu_{\mathbf{x}}\mu_{\mathbf{x}}} \quad (23)$$

$$a = \frac{\sigma_{\mathbf{xy}}^2}{\sigma_{\mathbf{x}}^2}, \quad (24)$$

where $\sigma_{\mathbf{xy}}^2$ is the covariance of \mathbf{x} and \mathbf{y} and $\sigma_{\mathbf{x}}^2$ is the variance of \mathbf{x} .

ACKNOWLEDGMENTS

The authors would like to thank Gregg Trahey and Jeremy Dahl for their feedback in the preparation of this work.

REFERENCES

- [1] R. S. C. Cobbold, *Foundations of Biomedical Ultrasound* (Biomedical Engineering). London, U.K.: Oxford Univ. Press, 2007, pp. 429–437.
- [2] H. Ahman, L. Thompson, A. Swarbrick, and J. Woodward, "Understanding the advanced signal processing technique of real-time adaptive filters," *J. Diagnostic Med. Sonography*, vol. 25, no. 3, pp. 145–160, 2009.
- [3] E. Boni, A. C. H. Yu, S. Freear, J. A. Jensen, and P. Tortoli, "Ultrasound open platforms for next-generation imaging technique development," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 7, pp. 1078–1092, Jul. 2018.
- [4] D. Guenther and W. Walker, "Optimal apodization design for medical ultrasound using constrained least squares part I: Theory," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 2, pp. 332–342, Feb. 2007.
- [5] I. K. Holfort, F. Gran, and J. A. Jensen, "Broadband minimum variance beamforming for ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, pp. 314–325, Feb. 2009.
- [6] B. Byram, K. Dei, S. Member, J. Tierney, and D. Dumont, "A Model and Regularization Scheme for Ultrasonic Beamforming Clutter Reduction," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 11, pp. 1–14, 2015.

- [7] J. Shin and L. Huang, "Spatial prediction filtering of acoustic clutter and random noise in medical ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 396–406, Feb. 2017.
- [8] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.
- [9] L. L. Brickson, D. Hyun, and J. J. Dahl, "Reverberation noise suppression in the aperture domain Using 3D fully convolutional neural networks," in *Proc. IEEE Int. Ultrason. Symp.*, Oct. 2018, pp. 1–4.
- [10] S. Krishnan, K. W. Rigby, and M. O'Donnell, "Efficient parallel adaptive aberration correction," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 45, no. 3, pp. 691–703, May 1998.
- [11] K. W. Hollman, K. W. Rigby, and M. O'Donnell, "Coherence factor of speckle from a multi-row probe," in *Proc. IEEE Ultrason. Symp.*, 1999, pp. 1257–1260.
- [12] P.-C. Li and M.-L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003.
- [13] J. Camacho, M. Parrilla, and C. Fritsch, "Phase coherence imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 5, pp. 958–974, May 2009.
- [14] W. Long, N. Bottenus, and G. E. Trahey, "Incoherent clutter suppression using lag-one coherence," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 8, pp. 1544–1557, Aug. 2020.
- [15] D. Shattuck, "Compound scanning with a phased array," *Ultrason. Imag.*, vol. 4, no. 2, pp. 93–107, Apr. 1982.
- [16] C. H. Seo and J. T. Yen, "Sidelobe suppression in ultrasound imaging using dual apodization with cross-correlation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 55, pp. 2198–2210, Oct. 2008.
- [17] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Short-lag spatial coherence of backscattered echoes: Imaging characteristics," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, pp. 1377–1388, Jul. 2011.
- [18] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes, "The delay multiply and sum beamforming algorithm in ultrasound B-Mode medical imaging," *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 940–949, Apr. 2015.
- [19] M. R. Morgan, G. E. Trahey, and W. F. Walker, "Multi-covariate imaging of sub-resolution targets," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1690–1700, Jul. 2019.
- [20] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 898–910, May 2019.
- [21] O. M. H. Rindal, A. Austeng, A. Fatemi, and A. Rodriguez-Molares, "The effect of dynamic range alterations in the estimation of contrast," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 7, pp. 1198–1208, Jul. 2019.
- [22] A. Rodriguez-Molares, O. Marius, H. Rindal, and D. Jan, "The Generalized Contrast-to-Noise ratio : A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2019.
- [23] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-Scans," *IEEE Trans. Sonics Ultrason.*, vol. 30, no. 3, pp. 156–163, May 1983.
- [24] M. Patterson, "The improvement and quantitative assessment of B-mode images produced by an annular array/cone hybrid," *Ultrason. Imag.*, vol. 5, no. 3, pp. 195–213, Jul. 1983.
- [25] S. W. Smith, R. F. Wagner, J. F. M. F. Sandrik, and H. Lopez, "Low contrast detectability and contrast/detail analysis in medical ultrasound," *IEEE Trans. Sonics Ultrason.*, vol. 3, no. 3, pp. 164–173, 1983.
- [26] R. J. Zemp, M. D. Parry, C. K. Abbey, and M. F. Insana, "Detection performance theory for ultrasound imaging systems," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 300–310, Mar. 2005.
- [27] K. Ranganathan and W. F. Walker, "Cystic resolution: A performance metric for ultrasound imaging systems," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, pp. 782–792, Apr. 2007.
- [28] D. A. Guenther and W. F. Walker, "Generalized cystic resolution: A metric for assessing the fundamental limits on beamformer performance," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 1, pp. 77–90, Jan. 2009.
- [29] K. Dei, A. Luchies, and B. Byram, "Contrast ratio dynamic range: A new beamformer performance metric," *IEEE Int. Ultrason. Symp.*, 2017, pp. 1–3.
- [30] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2008.
- [31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 1, p. 14, Dec. 2004.
- [32] V. Vo, "Fast algorithms for histogram matching: Application to texture synthesis," *J. Electron. Imag.*, vol. 9, no. 1, p. 39, Jan. 2000.
- [33] D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1143–1152, May 2006.
- [34] S. M. Pizer *et al.*, "Adaptive Histogram Equalization and Its Variants," *Comput. Vis., Graph., Image Process.*, vol. 39, pp. 355–368, Dec. 1987.
- [35] M. Karaman, P.-C. Li, and M. O'Donnell, "Synthetic aperture imaging for small scale systems," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 42, no. 3, pp. 429–442, May 1995.
- [36] G. E. Trahey, S. W. Smith, and O. T. von Ramm, "Speckle pattern correlation with lateral aperture translation: Experimental results and implications for spatial compounding," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 33, no. 3, pp. 257–264, May 1986.
- [37] J. J. Dahl, D. Hyun, M. Lediju, and G. E. Trahey, "Lesion detectability in diagnostic ultrasound with short-lag spatial coherence imaging," *Ultrason. Imag.*, vol. 33, no. 2, pp. 119–133, Apr. 2011.
- [38] N. Bottenus, B. C. Byram, J. J. Dahl, and G. E. Trahey, "Synthetic aperture focusing for short-lag spatial coherence imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, no. 9, pp. 1816–1826, Sep. 2013.
- [39] J. Thijssen, "Gray level transforms and lesion detectability in echographic images," *Ultrason. Imag.*, vol. 10, no. 3, pp. 171–195, Jul. 1988.
- [40] O. Huang *et al.*, "MimickNet, mimicking clinical image Post-processing under black-box constraints," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2277–2286, Jun. 2020.
- [41] C. Sparrow, "On spectroscopic resolving power," *Astrophys. J.*, vol. 44, pp. 76–86, Dec. 1916.



Nick Bottenus (Member, IEEE) received the B.S.E. degree in biomedical engineering and electrical and computer engineering and the Ph.D. degree in biomedical engineering from Duke University, Durham, NC, USA, in 2011 and 2017, respectively.

From 2017 to 2019, he was a Research Scientist with Duke University. He is currently an Assistant Professor with the Department of Mechanical Engineering, University of Colorado at Boulder, Boulder, CO, USA. His research interests include developing methods for performing large aperture ultrasound imaging and improving image quality through beamforming.



Brett C. Byram (Member, IEEE) received the B.S. degree in biomedical engineering and math from Vanderbilt University, Nashville, TN, USA, in 2004, and the Ph.D. degree in biomedical engineering from Duke University, Durham, NC, USA, in 2011.

He was a Research Assistant Professor with Duke University. In 2013, he joined Biomedical Engineering Department, Vanderbilt University, as an Assistant Professor. He has spent time working at Jørgen Jensen's Center for Fast Ultrasound, Lyngby, Denmark, and Siemens Healthcare's Ultrasound Division, Mountain View, CA, USA. He currently runs the Biomedical Elasticity and Acoustic Measurement (BEAM) Laboratory, where he and others in the lab pursue solutions to clinical problems using ultrasound. He is also with the Vanderbilt Institute for Surgery and Engineering (VISE) and the Vanderbilt University Institute of Imaging Science (VUIIS). His research interests include beamforming, motion estimation, and other related signal processing and hardware development tasks.



Dongwoon Hyun (Member, IEEE) received the B.S.E. and Ph.D. degrees in biomedical engineering from Duke University, Durham, NC, USA, in 2010 and 2017, respectively.

He is currently a Research Engineer with the Department of Radiology, Stanford University, Stanford, CA, USA. His current research interests include machine learning methods for beamforming, ultrasound molecular imaging, real-time software beamforming, and image quality assessment.