Information-Theoretic Feature Detection in Ultrasound Images

Greg Slabaugh, Gozde Unal, Ti-Chiun Chang
Siemens Corporate Research
Princeton, NJ 08540
{greg.slabaugh|gozde.unal|ti-chiun.chang}@siemens.com

Abstract

The detection of image features is an essential component of medical image processing, and has wide-ranging applications including adaptive filtering, segmentation, and registration. In this paper, we present an information-theoretic approach to feature detection in ultrasound images. Ultrasound images are corrupted by speckle noise, which is a disruptive random pattern that obscures the features of interest. Using theoretical probability density functions of the speckle intensity distributions, we derive analytic expressions that measure the distance between distributions taken from different regions in an ultrasound image and use these distances to detect features. We compare the technique to classic gradient-based feature detection methods.

1. INTRODUCTION

Ultrasound is one of the most commonly used medical imaging modalities. Compared to other modalities such as X-ray, MR, and PET, ultrasound scanning has many advantages, as it is fast, portable, relatively low cost, and presents virtually no risk to the patient.

However, the primary limitation of ultrasound is image quality. Ultrasound images are corrupted by speckle noise, an interference pattern resulting from the coherent accumulation of random scattering in a resolution cell of the ultrasound beam. While the texture of the speckle does not correspond to any underlying structure, the local brightness of the speckle pattern is related to the local echogenicity of the underlying scatterers. The speckle has a detrimental effect on the image quality and interpretability, and renders the detection of features in an ultrasound image as a difficult problem.

This work focuses on detecting salient edges in ultrasound images. For edge detection, the most popular approaches, such as the Canny edge detector [3], are based on gradient operators. The detector in this paper is perhaps most closely related to more recent information-theoretic methods such as [1]. However, unlike this method, our detector is designed for specific use with ultrasound images as it is based on speckle noise models. It differs from other ultrasound feature detection methods like the "sticks" approach [2] as our detector uses Rayleigh or Fisher-Tippett models, is based on information theory, and has a very simple implementation.

In this paper, we derive analytic expressions using the J-divergence to characterize the difference between ultrasound image regions that are modeled using Rayleigh or Fisher-Tippett distributions, which have been derived in the literature for modeling image intensities in speckled images. By comparing adjacent speckled regions in various directions (such as horizontally an vertically), we form a detector that identifies salient features embedded in speckle. We demonstrate the usefulness of our feature detector and compare its performance to other well-known operators, including the derivative of Gaussian, Canny edge detection, and non-parametric edge detection.

2. STATISTICAL MODELING OF ULTRASOUND SPECKLE

Figure 1 gives an overview of the ultrasound image formation process. After the RF data from the transducer is demodulated, one obtains an complex in-phase/quadrature image, $Q_I(x,y)$. In the case of fully formed speckle, which is typically assumed when the number of scatters per cell is greater than ten [4], it has been shown that the speckle in $Q_I(x,y)$ has a complex Gaussian distribution,

$$p(Q_I(x,y)) = \frac{1}{2\pi\sigma^2} e^{-|Q_I(x,y)|^2/(2\sigma^2)} \qquad (1)$$

where $Q_I(x,y)$ is complex. To produce a real image, envelope detection is performed by taking the magnitude of $Q_I(x,y)$. It is can be shown that under this transfor-

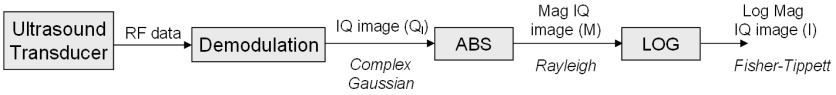


Figure 1. Block diagram of the ultrasound image formation process.

mation, the distribution in this magnitude image $M(x,y)$ becomes Rayleigh [5], i.e.,

$$p(M(x,y)) = \frac{M(x,y)}{\sigma^2} e^{-M(x,y)^2/(2\sigma^2)}, \tag{2}$$

where $M(x,y)$ is real. However, since $M(x,y)$ has a large dynamic range, it is customary to logarithmically transform the image to produce an image $I(x,y)$ suitable for display. Under this transformation, the fully formed speckle follows a Fisher-Tippett (FT) distribution,

$$p(I(x,y)) = 2e^{\left(2I(x,y) - \ln 2\sigma^2 - e^{2I(x,y) - \ln 2\sigma^2}\right)}. \tag{3}$$

More discussion and derivations of these distributions can be found in [4, 5, 6, 7].

2.1. Maximum Likelihood Rayleigh Estimator

Given a region Ω in $M(x,y)$, we would like to fit the data to the Rayleigh distribution. To proceed, we write the log likelihood of Equation 2 as $\ell(M(x,y),\sigma) = \ln\left(\int_\Omega p(M(x,y))dxdy\right)$,

$$\ell(M(x,y),\sigma) = \int_\Omega \left(\ln M(x,y) - \ln\sigma^2 - \frac{M(x,y)^2}{2\sigma^2}\right) dxdy. \tag{4}$$

Next, we can differentiate $\ell(M(x,y),\sigma)$ with respect to σ, and set this expression to zero to determine the maximum likelihood estimate of σ^2,

$$\sigma^2 = \frac{\int_\Omega M(x,y)^2 dxdy}{2\int_\Omega dxdy}. \tag{5}$$

Thus, given a region Ω, we can compute the maximum likelihood value of the parameter σ^2 from the image intensities in the region assuming the Rayleigh distribution.

2.2. Maximum likelihood Fisher-Tippett estimator

Similarly, by forming the log likelihood of Equation 3, we find an expression for σ^2 that is the maximum likelihood estimator of the FT distribution,

$$\sigma^2 = \frac{\int_\Omega e^{2I(x,y)} dxdy}{2\int_\Omega dxdy}. \tag{6}$$

3. INFORMATION-THEORETIC MATCHING OF REGIONS IN ULTRASOUND IMAGES

In this section we derive analytic expressions for measuring the distance between two distributions, p and q, taken from different windows of the image. Later, we will use these expressions in our feature detector.

3.1. Kullback-Liebler Divergence and J-Divergence

The Kullback-Liebler divergence, or relative entropy [8], is an information-theoretic measure between two distributions. The relative entropy $D(p||q)$ measures the inefficiency of assuming that a distribution is q when the true distribution is p. The Kullback-Liebler (KL) divergence is defined as

$$D(p||q) = \int p(x) \ln \frac{p(x)}{q(x)} dx. \tag{7}$$

In this definition, we follow the convention of defining $0\ln\frac{0}{0} = 0$ and $p\ln\frac{p}{0} = 0$. It is well-known that the KL divergence is asymmetric, that is, $D(p||q) \neq D(q||p)$. However, one can symmetrize the KL divergence using the J-divergence, $J = \frac{D(p||q)+D(q||p)}{2}$. It is useful to think of the J-divergence as a measure of the distance between two probability distributions, p and q.

3.2. Derivation of Rayleigh Case

For an image that can be modeled locally with Rayleigh distributions, we form a distribution p in one window of pixels, and another distribution q in another window of pixels, and model each window with a Rayleigh distribution, i.e.,

$$p(M(x,y)) = \frac{M(x,y)}{\sigma_1^2} e^{-M(x,y)^2/(2\sigma_1^2)} \tag{8}$$

$$q(M(x,y)) = \frac{M(x,y)}{\sigma_2^2} e^{-M(x,y)^2/(2\sigma_2^2)}, \tag{9}$$

where $M(x,y)$ is the intensity at pixel (x,y) in the magnitude IQ image and σ_1^2 and σ_2^2 are the parameter of each respective distribution. Then, we would like to

compute the J-divergence between the two distributions as a measure of how "different" the regions are. In the derivation below, we replace $M(x,y)$ in these expressions with x for simplicity. Furthermore, we derive the expression for $D(p||q)$ from which we can determine $D(q||p)$ by symmetry to get J:

$$D(p||q) = \int_0^\infty \frac{x}{\sigma_1^2} e^{-x^2/(2\sigma_1^2)} \ln \frac{\frac{x}{\sigma_1^2} e^{-x^2/(2\sigma_1^2)}}{\frac{x}{\sigma_2^2} e^{-x^2/(2\sigma_2^2)}} dx \quad (10)$$

Expanding the ln term yields

$$\begin{aligned} D(p||q) &= \int_0^\infty \frac{x}{\sigma_1^2} e^{-x^2/(2\sigma_1^2)} \left(\ln \frac{\sigma_2^2}{\sigma_1^2} - \frac{x^2}{2\sigma_1^2} + \frac{x^2}{2\sigma_2^2} \right) dx \\ &= \ln \left(\frac{\sigma_2^2}{\sigma_1^2} \right) \int_0^\infty \frac{x}{\sigma_1^2} e^{-x^2/(2\sigma_1^2)} dx \\ &\quad - \int_0^\infty \frac{x^3}{2\sigma_1^4} e^{-x^2/(2\sigma_1^2)} dx \\ &\quad + \int_0^\infty \frac{x^3}{2\sigma_1^2\sigma_2^2} e^{-x^2/(2\sigma_1^2)} dx, \end{aligned} \quad (11)$$

which, after some mathematics, gives

$$D(p||q) = \ln \left(\frac{\sigma_2^2}{\sigma_1^2} \right) - 1 + \frac{\sigma_1^2}{\sigma_2^2}. \quad (12)$$

Therefore, the J-divergence is then

$$J = -1 + \frac{\sigma_1^2}{2\sigma_2^2} + \frac{\sigma_2^2}{2\sigma_1^2}, \quad (13)$$

where σ_1^2 and σ_2^2 are determined from Equation 2.

3.3. Derivation of Fisher-Tippett Case

In the Fisher-Tippett case, we model regions p and q as Fisher-Tippett distributed regions with different parameters, σ_1^2 and σ_2^2. We derive an analytic expression for the Kullback-Liebler divergence of two regions described by Fisher-Tippett distributions, as

$$\begin{aligned} D(p||q) &= \int_0^\infty 2e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} \cdot \\ &\quad \ln \left(\frac{2e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}}}{2e^{2x-\ln 2\sigma_2^2 - e^{2x-\ln 2\sigma_2^2}}} \right) dx \end{aligned} \quad (14)$$

Expanding the ln term gives

$$\begin{aligned} D(p||q) &= 4\int_0^\infty x e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} dx \\ &\quad -2\ln 2\sigma_1^2 \int_0^\infty e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} dx \\ &\quad -2\int_0^\infty e^{4x-2\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} dx \\ &\quad -4\int_0^\infty x e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} dx \\ &\quad +2\ln 2\sigma_2^2 \int_0^\infty e^{2x-\ln 2\sigma_1^2 - e^{2x-\ln 2\sigma_1^2}} dx \\ &\quad +2\int_0^\infty e^{4x-\ln 2\sigma_1^2 - \ln 2\sigma_2^2 - e^{2x-\ln 2\sigma_1^2}} dx, \end{aligned} \quad (15)$$

which, after some mathematics, gives

$$\begin{aligned} D(p||q) &= e^{-\frac{1}{2\sigma_1^2}} \left(-\ln 2\sigma_1^2 + \ln 2\sigma_2^2 - 1 \right. \\ &\quad \left. -\frac{1}{2\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + \frac{1}{2\sigma_2^2} \right). \end{aligned} \quad (16)$$

The J-divergence is then

$$\begin{aligned} J &= \frac{1}{2} e^{-\frac{1}{2\sigma_1^2}} \left(-\ln 2\sigma_1^2 + \ln 2\sigma_2^2 - 1 \right. \\ &\quad \left. -\frac{1}{2\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + \frac{1}{2\sigma_2^2} \right) \\ &\quad + \frac{1}{2} e^{-\frac{1}{2\sigma_2^2}} \left(-\ln 2\sigma_2^2 + \ln 2\sigma_1^2 - 1 \right. \\ &\quad \left. -\frac{1}{2\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2\sigma_1^2} \right). \end{aligned} \quad (17)$$

where σ_1^2 and σ_2^2 are determined from Equation 3.

4. FEATURE DETECTION

At this point, we have tools in place to optimally estimate Rayleigh and Fisher-Tippett distributions in a region of an image. Furthermore, given two regions, we have an analytic expression for the distance between the distributions based on the J-divergence. In this section, we describe how these tools can be used for feature detection in ultrasound images.

4.1. Gradient-like Operator

One of the most commonly used methods to detect features in an image is the image gradient, computed via convolution of the image with a bandpass kernel, which is often modeled as the derivative of a Gaussian function. For example, the derivative kernel $K_x(x,y)$ in the x-dimension is

$$K_x(x,y) = \frac{\partial}{\partial x} \left(\frac{1}{2\pi\sigma_x\sigma_y} e^{-x^2/(2\sigma_x^2)} e^{-y^2/(2\sigma_y^2)} \right) \quad (18)$$

$$= \frac{-x}{2\pi\sigma_x^3\sigma_y} e^{-x^2/(2\sigma_x^2)} e^{-y^2/(2\sigma_y^2)} \quad (19)$$

where σ_x^2 and σ_y^2 are the variance in the x and y dimensions, respectively. Similarly, a kernel $K_y(x,y) =$

$K_x(x,y)^T$ can be found for the y dimension. The gradient can then be determined from $G_x(x,y) = K_x(x,y) * I(x,y)$, $G_y(x,y) = K_y(x,y) * I(x,y)$, where $I(x,y)$ is the image. The feature map is then simply the gradient magnitude, $F = \sqrt{G_x^2 + G_y^2}$. In Figure 2 (b), we show this gradient magnitude operator for a cardiac ultrasound image, for $\sigma_x = \sigma_y = \sigma = 2.5$. The gradient is sensitive to the speckle, which causes significant clutter in the feature map. While increasing the variance helps "blur over" the speckle, the effect of the speckle is still apparent in the feature map, and furthermore, larger values of σ blur detected edges, resulting in poorer localization.

In contrast, in our approach we use sliding windows, which are placed on either side of a pixel, as shown for two windows w_1 and w_2 in Figure 2 (a). We apply our FT model to $I(x,y)$, and our Rayleigh model to $M(x,y)$; here we describe the FT case. Given the set of pixels in w_1, we determine a FT parameter σ_1^2 using Equation 3, and likewise, we estimate σ_2^2 in w_2. Then, we compute J-divergence between these two distributions using Equation 17 as a measure of how different the regions are. When the windows are placed to the left and to the right of the pixel, this gives a horizontal distance map $J_x(x,y)$ that is functionally similar to the gradient operator in the x direction, except that the values are non-negative. This can be repeated in the y direction. Here, we define a feature map $F_j(x,y)$ as

$$F_j(x,y) = \sqrt{J_x(x,y)^2 + J_y(x,y)^2}. \qquad (20)$$

Figure 2 (c) shows an example of a cardiac ultrasound image and its feature map F_j. Note that this feature detector only picks up the most salient features and is much less distracted by the speckle compared to the gradient estimator.

The only real parameter to this feature detection method is the window size. Increasing the window size gives a better statistical modeling of the distribution's parameter in the window, and varies the scale of the features detected. For example, in Figure 3 we show the effect of changing the window size. We observe that the size of the features detected is proportional to the window size.

Another example is shown in Figure 4 for a lesion phantom. In (a), we show the original image. In (b), we show the J-divergence feature map applied to $M(x,y)$ using the Rayleigh distribution, and in (c) we show the J-divergence feature map applied to $I(x,y)$ using the Fisher-Tippett distribution, both using a window size of 7 by 7. Note that the results in (b) and (c) are nearly identical, which suggests our modeling is correct. In (d), we show the derivative of Gaussian feature detector, which was performed using $\sigma = 2.5$, for which the detected features have approximately the same size as those in (b) and (c). In (e), we show the output of the Canny operator, which was optimized to detect the salient edges while minimizing false detection of speckle edges. Finally, in (f) we show the output of using the J-divergence non-parametrically, i.e., computing $J = \frac{D(p||q)+D(q||p)}{2}$ for Parzen-windowed histograms. This latter method results in false detections due to speckle compared to the parametric methods. Compared with these other methods, our parametric feature detectors have a strong feature response while at the same time mitigating false responses due to the speckle.

5. CONCLUSION

In this paper, we presented an information theoretic approach to detect features in ultrasound images. Our feature detector is computed using the J-divergence of two Rayleigh or Fisher-Tippett distributed variables estimated from windows in the image. We demonstrated the ability of our method to detect features in ultrasound images and compared to other common feature detectors. For future work, we are interested in fully validating the method and using it in applications such as filtering and segmentation.

References

[1] A. B. Hamza, "Nonextensive information-theoretic measure for image edge detection," *Journal of Electronic Imaging*, vol. 15, no. 1, 2006.

[2] R. N. Czerwinski, D. L. Jones, and W. D. O'Brien, "Line and Boundary Detection in Speckle Images," *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1700–1714, 1998.

[3] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Patt. Anal. and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

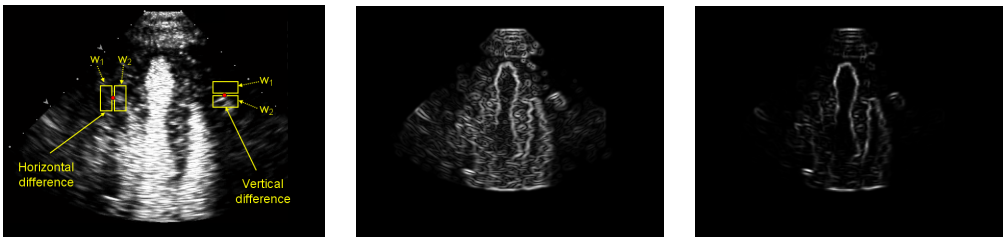
[4] V. Dutt and J. Greenleaf, "Statistics of the Log-Compressed Envelope," *Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3817–3825, 1996.

[5] J. Goodman, *Speckle Phenomena: Theory and Applications*, 1st ed. Work in Progress, 2005.

[6] R. Wagner, S. Smith, J. M. Sandrik, and H. Lopez, "Statistics of Speckle in Ultrasound B-Scans," *IEEE Transactions on Sonics and Ultrasonics*, vol. 30, no. 3, pp. 156–163, 1983.

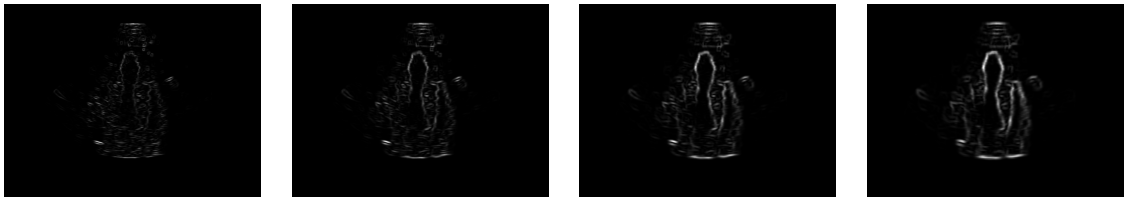
[7] O. Michailovich and D. Adam, "Robust Estimation of Ultrasound Pulses Using Outlier-Resistant De-Noising," *IEEE Trans. on Medical Imaging*, vol. 22, no. 3, pp. 368–392, 2003.

[8] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John-Wiley and Sons, 1991.



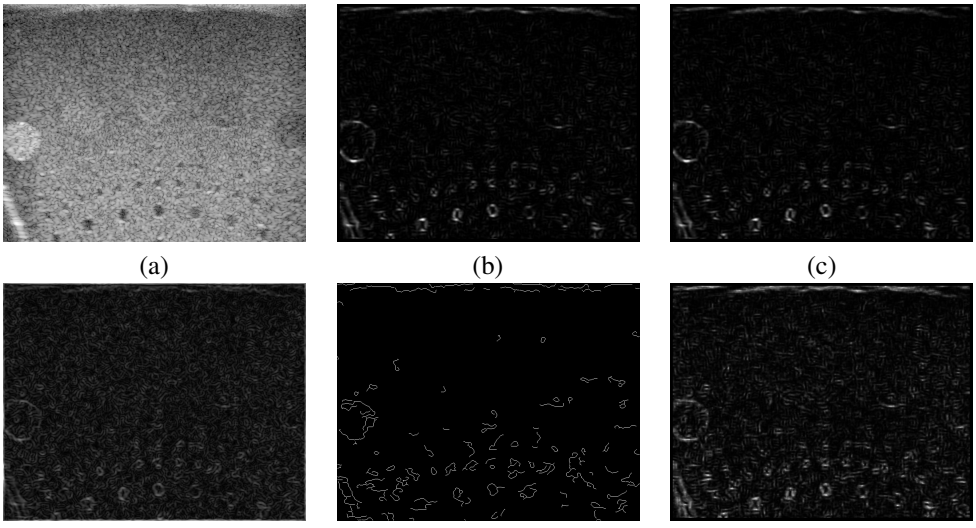
(a) (b) (c)

Figure 2. Gradient-like feature detection in a cardiac ultrasound image. Image (a), gradient (b), and J-divergence feature map (c) computed on the log magnitude IQ image using the Fisher-Tippett method. Please see the digital version of the images for maximal quality.



(a) (b) (c) (d)

Figure 3. Effect of window size. From left to right: 3x3, 5x5, 7x7, 9x9.



(a) (b) (c)

(d) (e) (f)

Figure 4. Different ultrasound feature detectors on a lesion phantom image. Original image (a), Rayleigh (b), Fisher-Tippett (c), Derivative of Gaussian (d), Canny (e), Non-Parametric (f).