A beamformer-independent method to predict photoacoustic visual servoing system failure from a single image frame

Eduardo A. Gonzalez*, Fabrizio Assis†, Jonathan Chrispin†, and Muyinatu A. Lediju Bell*‡§
*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD
†Division of Cardiology, Johns Hopkins Medical Institutions, Baltimore, USA
‡Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD
§Department of Computer Science, Johns Hopkins University, Baltimore, MD

2021 IEEE International Ultrasonics Symposium (IUS) | 978-1-6654-0355-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/IUS52206.2021.9593672

Abstract*—Visual servoing is a robotic method that has the potential to assist surgeons with tracking tool tips when attached to optical fibers to create photoacoustic images that are autonomously monitored. Currently, this approach must be tested with multiple image frames and multiple laser energies prior to each surgery in order to identify the minimum required energy that will not cause system failure over the number of frames tested. This study investigates possible integration of the generalized contrast-to-noise ratio (gCNR) into pre-surgical procedures as a method to predict system failure from only a single image frame. Photoacoustic data were acquired from an optical fiber inserted in a plastisol phantom or in the left ventricle of an *in vivo* swine heart. Raw data were processed with delay-and-sum (DAS) and short-lag spatial coherence (SLSC) beamforming (M = 25). gCNR values were estimated from a 3 mm x 3 mm region of interest (ROI) surrounding the photoacoustic target coordinates provided by the visual servoing algorithm. The prediction function modelled from phantom data was fit with R^2 values of 0.992 and 0.991 for DAS and SLSC beamformers, respectively. When applying this fit to the *in vivo* test data, the RMSE between measured segmentation accuracy and the prediction functions was 9.34% for DAS images and 4.78% for SLSC images. These results indicate that the newly introduced image quality metric gCNR has sufficient robustness to predict the performance of visual servoing segmentation tasks and thereby mitigate the burden, time, and requirements of testing multiple image frames prior to the initiation of a surgery.

I. Introduction

Visual servoing [1] is a promising approach to maintain visualization of surgical tool tips and nearby anatomical targets during minimally invasive procedures. The approach broadly refers to vision-based robot control. The robot "vision" that we focus on in this paper is provided through photoacoustic images [2], [3].

Photoacoustic imaging is achieved by transmitting pulsed light to a structure of interest, which absorbs the light, undergoes thermal expansion, and generates an acoustic response that is received by a conventional ultrasound probe [4]. This photoacoustic imaging technique was previously demonstrated for multiple applications that require surgery or interventions [5], such as visualization of brachytherapy seeds [6], [7], intravascular imaging [8], cardiac catheter visualization [3], and endonasal surgeries [9], [10]. In these applications, structures of interest include blood vessels, nerves, drill tips, and catheter or needle tips [3], [11]. In addition, one or more optical fibers may be coupled to the tool, catheter, or needle tips in order to transmit the light pulses [11], [12].

For image-guided surgery procedures, photoacoustic-based visual servoing involves three main steps. First, a photoacoustic signal is created, then processed in real time with a conventional delay-and-sum beamformer or with an advanced beamforming technique [13]. Second, the photoacoustic image is processed by a segmentation algorithm in order to identify patterns or a specific target in a region of interest, such as the tip of an optical fiber attached to a cardiac catheter [3]. Finally, the robot moves the ultrasound probe to laterally center the target in the displayed image. This movement is applied based on the difference of the estimated centroid of the segmented mask and the lateral center of the imaging plane.

The success of the segmentation task relies on the photoacoustic image quality, which can often be enhanced by increasing the incident laser energy. On the other hand, increasing the laser energy must be performed within safety limits. To strike an appropriate balance between these two competing interests [14], a series of calibration tests may be conducted prior to surgery to assess segmentation performance at multiple laser energies. Depending on the number of energy levels tested before identifying an optimal value, this calibration step can be time-consuming. In addition, consecutive failure events will trigger a regional search task that may cause the ultrasound probe placement to deviate from its target position [13]. Therefore, a faster calibration method is required, which can be achieved by necessitating only one image to make predictions about segmentation accuracy for visual servoing.

This study investigates the integration of the generalized contrast-to-noise ratio (gCNR) [15], [16] into pre-surgical procedures as a method to predict system failure of visual servoing tasks from the analysis of a single image frame. We created models of segmentation accuracy as a function of image quality (based on calibrated phantom data), referred to as *segmentation prediction curves*. These predictions were then compared with the segmentation accuracy obtained with

978-0-7381-1209-1/21/$31.00 ©2021 IEEE

corollary data obtained from an *in vivo* swine heart. Segmentation prediction curves were first created using the image quality metric of gCNR, then compared with results achieved with signal-to-noise ratio (SNR), contrast, and contrast-to-noise ratio (CNR) as alternative image quality metrics.

II. Method

A. Data acquisition

The photoacoustic imaging equipment consisted of a 1-mm-diameter optical fiber connected to a Phocus Mobile laser (Opotek Inc., Carlsbad, CA, USA). The laser was synchronized to receive signals from an E-CUBE 12R research ultrasound system (Alpinion, Seoul, South Korea) connected to either an L3-8 linear array ultrasound probe or a SP1-5 phased array ultrasound probe (Alpinion, Seoul, South Korea). When using this system to acquire data from the following phantom and *in vivo* experiments, 50 frames were acquired per reported laser energy.

To perform phantom experiments, the free end of the optical fiber was inserted into a plastisol phantom at 24 mm depth from the L3-8 ultrasound probe. To ensure that the tip of the optical fiber was positioned within the imaging plane, a laser energy of 1.00 mJ was set, and the ultrasound probe was translated across the elevation axis until the amplitude of the photoacoustic signal was maximized. After the ultrasound transducer was fixed in position, photoacoustic data was acquired with the fiber transmitting laser light with a wavelength of 750 nm, with laser energies of 13.5, 43, 110, 120, 805, 1040, 1241, 1433, 1690, 1760, 1823, and 2330 μJ.

To perform *in vivo* cardiac experiments, a 1-mm-diameter optical fiber was inserted within a cardiac catheter, which was inserted in the femoral vein of an *in vivo* swine. The catheter was then navigated through the inferior vena cava toward the heart, with the ventricular septum as the final destination. The fiber was located at 63 mm depth from the SP1-5 ultrasound probe. Photoacoustic data were acquired with a laser wavelength of 750 nm, with laser energies of 2.9, 6.4, 12.3, 18.7, 24.9, 31.0, 37.2, 43.3, 49.4, 55.6, 67.9, 80.2, 92.5, 104.7, 117.0, and 608.5 μJ. This *in vivo* procedure was approved by the Johns Hopkins University Animal Care and Use Committee.

B. Estimation of segmentation accuracy

Our framework for calculating the segmentation accuracy of a visual servoing task is illustrated in Fig. 1. Raw data from each energy were processed with delay-and-sum (DAS) and short-lag spatial coherence (SLSC) beamforming (M = 25) [17], both displayed with 15-dB dynamic range. The visual servoing algorithm [3] then estimated the centroid of the identified photoacoustic signal (i.e., tip of the optical fiber) and stored the segmentation success or failure as a binary flag. Following this step, gCNR [15], [16] values were estimated from a 3 mm $\times$ 3 mm region of interest (ROI) surrounding the target coordinates. Finally, a histogram analysis of the successful and failed segmentation events was conducted over 50 bins of gCNR values, where the segmentation accuracy γ for each gCNR bin number was determined using the equation:

$$\gamma(\text{gCNR bin \#}) = \frac{\#\text{Success}_{\text{gCNR}}}{\#\text{Success}_{\text{gCNR}} + \#\text{Failure}_{\text{gCNR}}} \times 100\%. \quad (1)$$

To compare the segmentation accuracy between phantom and *in vivo* experiments, a segmentation prediction curve was first generated by fitting a sigmoid function to γ measurements from the the phantom experiment. The prediction error was then calculated by computing the RMSE between the segmentation prediction curve and the γ obtained from the *in vivo* experiment. For comparison, the method for calculating γ in Eq. (1) was investigated by substituting gCNR with an alternative image quality metric (i.e., SNR, contrast, or CNR).

C. Image quality metrics

The following definitions were used to calculate gCNR, SNR, contrast, and CNR:

$$\text{gCNR} = 1 - \sum_{k=0}^{N-1} \min\{h_i(x_k), h_o(x_k))\} \quad (2)$$

$$\text{SNR} = \mu_i/\sigma_o \quad (3)$$

$$\text{Contrast} = \mu_i - \mu_o \quad (4)$$

$$\text{CNR} = |\mu_i - \mu_o|/\sqrt{\sigma_i^2 + \sigma_o^2}, \quad (5)$$

where $h(x_k)$ is a histogram evaluated at bin k centered on x_k, μ and σ represent the mean and standard deviation, respectively, of signals within regions located inside and outside the target, as represented by the subscripts i and o, respectively. These metrics were evaluated on uncompressed envelope-detected DAS images for and uncompressed SLSC images. A total of 100 bins were used to create the histograms for gCNR measurements.

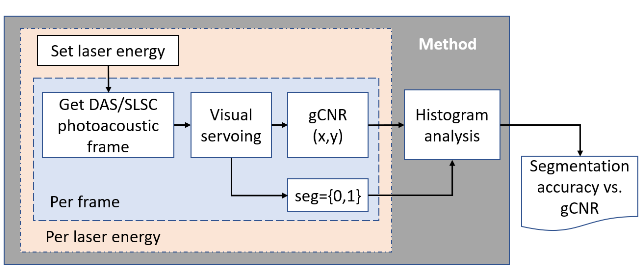

Fig. 1: Framework for calculating the segmentation accuracy of visual servoing tasks through a histogram analysis of gCNR

III. Results and Discussion

Fig. 2 shows examples of ultrasound images from the *in vivo* swine heart, containing overlays of photoacoustic images created with different beamformers and different laser energies. With the lower laser energy (top row), the example SLSC photoacoustic image shows improved visualization of the target, producing a gCNR value of 0.96 in comparison to 0.49 gCNR value produced with DAS, which resulted in a failed segmentation. With the higher laser energy (bottom row), the gCNR values obtained with DAS and SLSC were

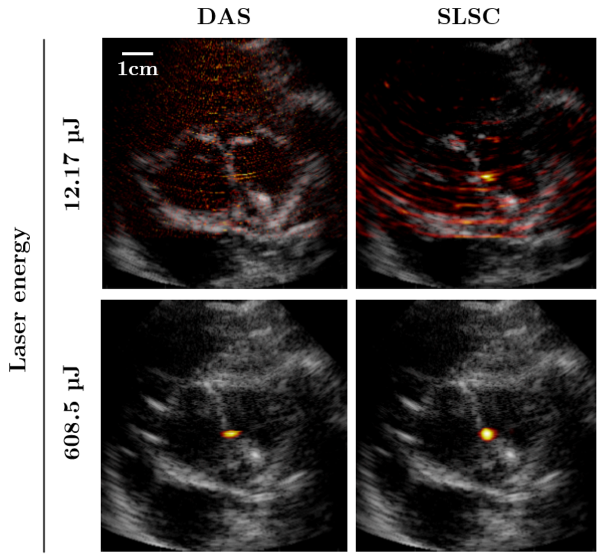

Fig. 2: Examples of photoacoustic images from an *in vivo* swine heart, processed with DAS and SLSC beamformers, obtained with laser energies of 12.17 μJ and 608.5 μJ. Photoacoustic images are overlaid on corresponding harmonic ultrasound images.

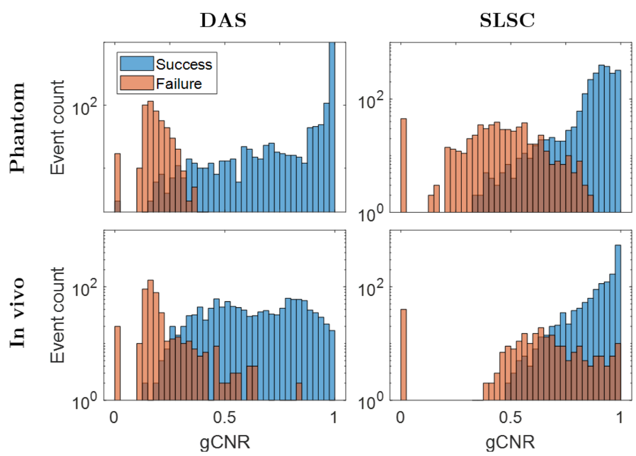

Fig. 3: Estimated histograms of successful and failed segmentation cases sorted by the gCNR value obtained with DAS and SLSC beamforming.

0.89 and 0.99, respectively, and both results produced successful segmentations while visual servoing based on these images.

Fig. 3 shows example histograms of successful and failed segmentation cases, sorted by the gCNR values obtained with DAS and SLSC beamforming. The bin at gCNR=0 represents failed segmentation events where no coordinates of the target were found after running the visual servoing algorithm. Overall, segmentation fails more often for low gCNR values, which are generally associated with lower laser energies. We define a 50% γ threshold at the gCNR bin value where there are similar counts of success and failure events (i.e., where the red and blue bars that overlap are of similar height).

For the *in vivo* experiments reported at the bottom of Fig. 3,

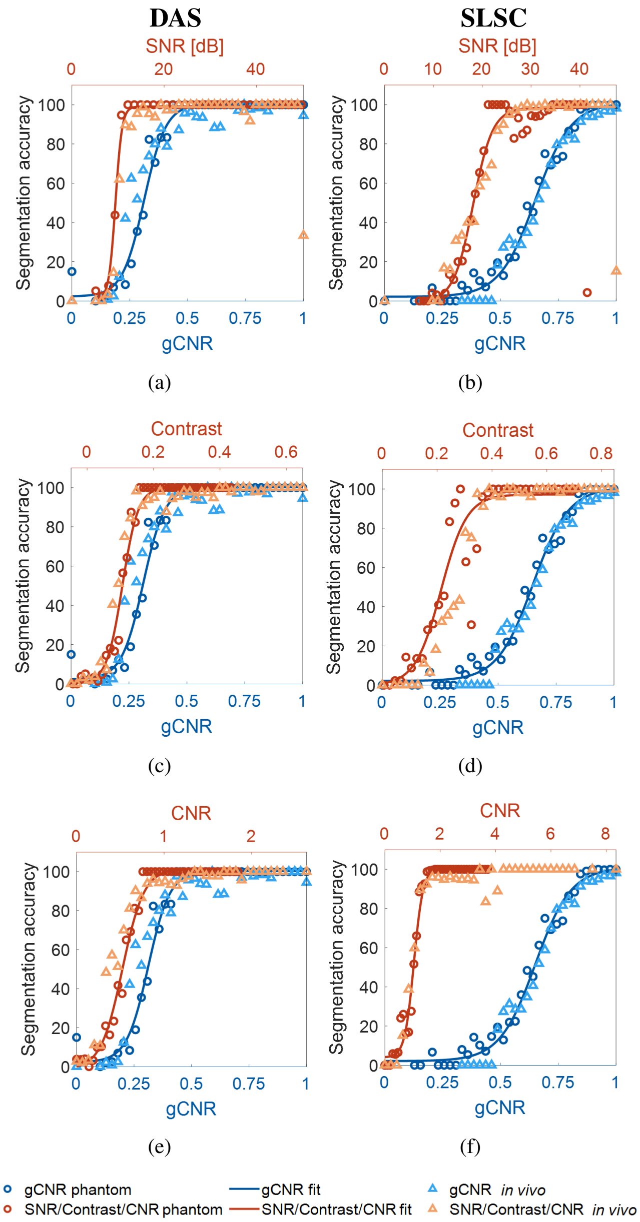

Fig. 4: Segmentation prediction curves with datapoints showing measured segmentation accuracy (i.e., γ) as functions of the gCNR, SNR, contrast, and CNR of (left) DAS and (right) SLSC images.

DAS and SLSC achieved 50% segmentation accuracy (i.e., 50% γ) at gCNR values of 0.288 and 0.663, respectively, likely because the signal region of the DAS image contains large amplitude variations, which are known to decrease the gCNR of photoacoustic images [15]. However, the visual servoing algorithm compensates for these large amplitude variations through various morphological operations [2]. Therefore, the DAS segmentation is considered successful despite the low gCNR value.

Fig. 4 shows the segmentation prediction curves modelled from phantom data when using gCNR as the image quality metric for the DAS ($R^2 = 0.992$) and SLSC ($R^2 = 0.991$)

beamformers. These data and predictions are replicated thrice per beamformer in order to compare with segmentation prediction curves when using SNR, contrast, and CNR as the image quality metrics. Predictions are additionally compared to the *in vivo* γ results (shown as triangles). With gCNR as the image quality metric, the RMSE between *in vivo* data and corresponding predictions was 9.34% for DAS images and 4.78% for SLSC images. A summary of the R^2 of each sigmoid fit in Fig. 4 and corresponding RMSE values between *in vivo* γ results and predictions are reported in Table I. Note that the gCNR metric showed comparable performance to the SNR and contrast metrics when predicting the segmentation accuracy from DAS images, which further highlights that our proposed approach using gCNR has the potential to be independent of the chosen beamformer for visual servoing.

Regarding interpretation of the segmentation prediction curves, one that contains a gradual (rather than steep) slope is generally more beneficial. This feature is desired because it enables more robust predictions in the presence of minor frame-to-frame variations that cause large variations for a particular image quality metric. Extending this concept to the results in Fig. 4, the gCNR metric applied to SLSC images contains the desired gradual changes more often. For example, when selecting regions of comparable width from the x-axes of Fig. 4, (e.g., 0.627-0.667 for gCNR and 0.863-1.197 for CNR in Fig. 4(f)), the rate of change in segmentation accuracy differs (e.g., gCNR results in a 14.78% change between the first and last data points of the selected range, while CNR results in a 44.78% change).

Overall, our results indicate that gCNR has sufficient robustness to predict the performance of visual servoing segmentation tasks and thereby mitigate the burden, time, and requirements of testing multiple image frames prior to the initiation of a surgery. While these predictions may be possible with alternative image quality metrics, the rate of change is generally steeper, with the slope fixed based on the minimum and maximum values measured with each metric.

TABLE I: Summary of R^2 values (for segmentation prediction curves) and RMSE (between predictions and *in vivo* results) for each beamformer and image quality metric in Fig. 4

Image Quality Metric	DAS R^2	DAS RMSE	SLSC R^2	SLSC RMSE
SNR	0.935	6.39%	0.961	38.32%
gCNR	0.992	9.34%	0.991	4.78%
Contrast	0.998	7.91%	0.914	13.14%
CNR	0.988	10.45%	0.985	8.16%

IV. Conclusions

This work is the first to integrate gCNR, visual servoing, and photoacoustic imaging. By modeling segmentation prediction curves from phantom data, gCNR successfully predicts the segmentation accuracy of *in vivo* photoacoustic DAS and SLSC images. Overall, the proposed method has sufficient robustness to predict the performance of visual servoing segmentation tasks and thereby alleviate the burden, time, and requirements of testing multiple image frames during prior to a surgery or interventional procedure.

Acknowledgments

This work was supported by NSF CAREER Award ECCS-1751522 and NSF SCH Award IIS-2014088. The authors acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] P. Abolmaesumi, S. E. Salcudean, W.-H. Zhu, M. R. Sirouspour, and S. P. DiMaio, "Image-guided control of a robot for medical ultrasound," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 1, pp. 11–23, 2002.

[2] M. A. L. Bell and J. Shubert, "Photoacoustic-based visual servoing of a needle tip," *Scientific Reports*, vol. 8, no. 1, p. 15519, 2018.

[3] M. Graham, F. Assis, D. Allman, A. Wiacek, E. Gonzalez, M. Gubbi, J. Dong, H. Hou, S. Beck, J. Chrispin, and M. A. L. Bell, "In vivo demonstration of photoacoustic image guidance and robotic visual servoing for cardiac catheter-based interventions," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1015–1029, 2020.

[4] P. Beard, "Biomedical photoacoustic imaging," *Interface Focus*, vol. 1, no. 4, pp. 602—-631, 2011.

[5] A. Wiacek and M. A. L. Bell, "Photoacoustic-guided surgery from head to toe," *Biomedical Optics Express*, vol. 12, no. 4, pp. 2079–2117, 2021.

[6] M. A. L. Bell, N. P. Kuo, D. Y. Song, J. U. Kang, and E. M. Boctor, "In vivo visualization of prostate brachytherapy seeds with photoacoustic imaging," *Journal of Biomedical Optics*, vol. 19, no. 12, p. 126011, 2014.

[7] K. L. Bell, T. Harrison, N. Usmani, and R. J. Zemp, "Integrated transrectal probe for translational ultrasound-photoacoustic imaging," in *Photons Plus Ultrasound: Imaging and Sensing 2016*, vol. 9708. International Society for Optics and Photonics, 2016, p. 97080A.

[8] A. B. Karpiouk, B. Wang, and S. Y. Emelianov, "Development of a catheter for combined intravascular ultrasound and photoacoustic imaging," *Review of Scientific Instruments*, vol. 81, no. 1, p. 014901, 2010.

[9] M. A. L. Bell, A. K. Ostrowski, K. Li, P. Kazanzides, and E. M. Boctor, "Localization of transcranial targets for photoacoustic-guided endonasal surgeries," *Photoacoustics*, vol. 3, no. 2, pp. 78–87, 2015.

[10] M. T. Graham, J. Huang, F. Creighton, and M. A. L. Bell, "Simulations and human cadaver head studies to identify optimal acoustic receiver locations for minimally invasive photoacoustic-guided neurosurgery," *Photoacoustics*, p. 100183, 2020.

[11] B. Eddins and M. A. L. Bell, "Design of a multifiber light delivery system for photoacoustic-guided surgery," *Journal of Biomedical Optics*, vol. 22, no. 4, p. 041011, 2017.

[12] D. Piras, C. Grijsen, P. Schutte, W. Steenbergen, and S. Manohar, "Photoacoustic needle: minimally invasive guidance to biopsy," *Journal of Biomedical Optics*, vol. 18, no. 7, p. 070502, 2013.

[13] E. A. Gonzalez and M. A. L. Bell, "GPU implementation of photoacoustic short-lag spatial coherence imaging for improved image-guided interventions," *Journal of Biomedical Optics*, vol. 25, no. 7, p. 077002, 2020.

[14] M. A. Lediju Bell, "Photoacoustic imaging for surgical guidance: Principles, applications, and outlook," *Journal of Applied Physics*, vol. 128, no. 6, p. 060904, 2020.

[15] K. M. Kempski, M. T. Graham, M. R. Gubbi, T. Palmer, and M. A. L. Bell, "Application of the generalized contrast-to-noise ratio to assess photoacoustic image quality," *Biomedical Optics Express*, vol. 11, no. 7, pp. 3684–3698, 2020.

[16] A. Rodriguez-Molares, O. M. H. Rindal, J. D'hooge, S.-E. Måsøy, A. Austeng, M. A. L. Bell, and H. Torp, "The generalized contrast-to-noise ratio: a formal definition for lesion detectability," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, 2019.

[17] M. T. Graham and M. A. L. Bell, "Photoacoustic spatial coherence theory and applications to coherence-based image contrast and resolution," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 10, pp. 2069–2084, 2020.