

Performance test methods for near-infrared fluorescence imaging

Udayakumar Kanniyappan

*Fischell Department of Bioengineering, University of Maryland, College Park, MD, USA
Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA*

Bohan Wang and Charles Yang

Fischell Department of Bioengineering, University of Maryland, College Park, MD, USA

Pejman Ghassemi

Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA

Maritoni Litorja

National Institute of Standards and Technology, Gaithersburg, MD, USA

Nitin Suresh

Fischell Department of Bioengineering, University of Maryland, College Park, MD, USA

Quanzeng Wang

Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA

Yu Chen^{a)}

*Fischell Department of Bioengineering, University of Maryland, College Park, MD, USA
Department of Biomedical Engineering, University of Massachusetts Amherst, MA, USA*

T. Joshua Pfefer^{a)}

Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA

(Received 21 October 2019; revised 3 March 2020; accepted for publication 10 April 2020; published 1 June 2020)

Purpose: Near-infrared fluorescence (NIRF) imaging using exogenous contrast has gained much attention as a technique for enhancing visualization of vasculature using untargeted agents, as well as for the detection and localization of cancer with targeted agents. In order to address the emerging need for standardization of NIRF imaging technologies, it is necessary to identify the best practices suitable for objective, quantitative testing of key image quality characteristics. Toward the development of a battery of test methods that are rigorous yet applicable to a wide variety of devices, we have evaluated techniques for phantom design, measurement, and calculation of specific performance metrics.

Methods: Using a NIRF imaging system for indocyanine green imaging, providing excitation at 780 nm and detection above 830 nm, we explored methods to evaluate uniformity, field of view, spectral crosstalk, spatial resolution, depth of field, sensitivity, linearity, and penetration depth. These measurements were performed using fluorophore-doped multiwell plate and high turbidity planar phantoms, as well as a 3D-printed multichannel phantom and a USAF 1951 resolution target.

Results and Conclusions: Based on a wide range of approaches described in medical and fluorescence imaging literature, we have developed and demonstrated a cohesive battery of test methods for evaluation of fluorescence image quality in wide-field imagers. We also propose a number of key metrics that can facilitate direct, quantitative comparison of device performance. These methods have the potential to facilitate more uniform evaluation and inter-comparison of clinical and preclinical imaging systems than is typically achieved, with the long-term goal of establishing international standards for fluorescence image quality assessment. © 2020 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.14189]

Key words: fluorescence imaging, indocyanine green, near-infrared, standards, test methods, tissue phantoms

1. INTRODUCTION

In the past decade, there have been major advances in fluorescence-based imaging techniques for medical diagnostics,

including exogenous near-infrared (NIR) fluorophores which enhance the information collected by these devices. Near-infrared excitation and emission wavelengths (690–1000 nm) represent a region where endogenous tissue fluorescence tends to be low and light penetration is relatively high, due to lower absorption of water, melanin, oxy, and de-oxy hemoglobin.^{1,2} *En-face*, or surface NIR fluorescence imaging with digital cameras has been implemented for a wide range of

[Correction added on 12 June 2020 after first online publication: The fourth author's first name has been changed from 'Pejman' to 'Pejman' in this version]

applications, such as metastatic imaging,³ lymph node identification,^{4–6} intraoperative tumor, delineation, and vascular mapping.⁷ While the development of NIR imaging exhibits tremendous potential for clinical improvements, there remains a lack of standardized test methods for objective, quantitative characterization of device performance.

Well-validated tissue-simulating phantoms can facilitate a wide variety of performance evaluation tasks throughout the device life cycle, including early system development, device optimization and inter-comparisons, clinical trial standardization, regulatory clearance, manufacturing quality control, recalibration, clinical constancy testing, and clinician training, among others. Currently, there are numerous international consensus documents that describe standardized phantom-based test methods for established medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and PET.^{8–10} However, no equivalent documents exist for optical imaging modalities such as NIR fluorescence (NIRF). Thus, there is a need to identify an optimal set of performance metrics that are objective, quantitative, and scientifically rigorous, yet minimally burdensome for users. While a standard set of test methods may not be appropriate for all NIR fluorescence devices, they can often serve to provide guiding principles that can be adapted for novel system designs.

Prior studies on medical imaging,^{8–10} NIRF,^{1–4,11–18} and other optical modalities (e.g., hyperspectral)^{19,20} identify key characteristics relevant for *en-face* imaging. Several key commonly cited metrics are spatial resolution,^{15,21–29} sensitivity/detectability,^{1,7,11–13,15–17,21–31} linearity,^{11,13,17,32,33} uniformity,^{16,24,34} depth of field (DOF),^{35–37} field of view (FOV),^{11–15} excitation light crosstalk,^{16,24,25,29,38,39} and penetration depth.^{16,24,26,27,29,39} Some metrics that are commonly used with white light images are essential in assessing the image quality of NIRF imaging systems, such as spatial resolution,^{15,21–29} DOF,^{35–37} and distortion.^{40,41} A recent study of fluorescence tomography image quality implemented phantoms with (a) indocyanine green (ICG)-doped objects in the 0.25–2 μM range at a single depth and (b) objects at a single concentration but a range of depths from 4 to 14 mm;¹⁸ this study also used signal-to-noise ratio (SNR) as a key variable. In a frequency-domain fluorescence study, detectability was evaluated as a function of inclusion and background fluorophore concentrations, as well as object size.³⁰ In spite of the wide range of techniques described in optical image quality literature and many contributions of these works, the methods described in NIRF surface imaging studies are typically or less comprehensive than necessary for standardization or lacking in rigor (e.g., sensitivity data without objective, quantitative evaluation of linearity, and detection limits).

Therefore, the purpose of this study was to develop a basic battery of performance test methods inspired by prior work in fluorescence and other imaging modalities. Specifically, we have generated tissue-simulating phantoms with clinically relevant fluorophores using indocyanine green (ICG) and used them to characterize the image quality of a custom NIRF

imaging system. In some cases, we have performed evaluations using multiple approaches and discuss the relative merits of each approach.

2. MATERIALS AND METHODS

2.A. NIRF imaging system

All images analyzed in this study were acquired with a custom benchtop NIRF imaging system (Fig. 1). A light-emitting diode (M780L3, Thorlabs, Inc., Newton, NJ) was used as the illumination light source, with 780 nm center wavelength and 30 nm bandwidth, and irradiance at the sample surface was 2 mW/cm². A convex lens and diffuser were used to increase the uniformity of illumination. An 800 nm short-pass filter (84–729, Edmund Optics, Barrington, NJ) was used to reduce the potential of detecting excitation light from the diode. A long-pass filter with a cut-off wavelength of 825 nm (86–078, Edmund Optics, Barrington, NJ) was secured to the camera. Fluorescence images were captured using a 16-bit CCD camera (Alta U2000, Apogee Imaging Systems, Roseville, CA) with the setup of a zoom lens (75 mm focal length, *f*/3.9, Tamron, Commack, NY). The CCD was mounted on a stage fitted with a 125-mm-long travel rack and pinion track to move the camera vertically. Camera images were obtained using Micro Manager software⁴² (μ manager v1.1) and ImageJ⁴³ was used for postprocessing. Depending on the image geometry and level of distortion, it is sometimes preferable to determine field of view (FOV) in terms of horizontal and vertical distances, or in terms of maximum angular extent. In our system, the FOV was 8.9 cm (7.3°) in the horizontal direction and 6.7 cm (5.4°) in the vertical direction.⁴⁴

2.B. Phantom fabrication

We developed three types of tissue-mimicking phantoms as components of the performance tests, using ICG as the fluorophore. ICG is a water-soluble dye that has broad absorption around 780 nm and an emission peak near 800 nm. Due to its optical properties and its biocompatibility, ICG has become a popular fluorophore for clinical imaging of biological structures such as vessels and ducts.^{1–7} Prior to fluorescence phantom fabrication, the material used in multiwell and wide-field phantoms was calibrated so as to simulate the scattering of tissue ($\mu_s' = 10 \text{ cm}^{-1}$). To achieve this, we constructed 1-mm-thick slab phantoms using epoxy-resin with varying concentrations of titanium dioxide (TiO₂). Each epoxy-resin (EasyCast, Environmental Technology Inc., Fields Landing, CA) phantom was prepared as per standard protocol from the manufacturer by mixing 1:1 (resin:hardener) ratio by weight. The mixture was then stirred for 10 min and kept at low pressure to remove any air bubbles. The prepared phantom then was allowed to cure for 24 h. The diffuse reflectance and transmittance of thin slabs were measured using a UV-VIS spectrophotometer (Lambda 1050, Perkin Elmer, Waltham, MA), and the optical properties of

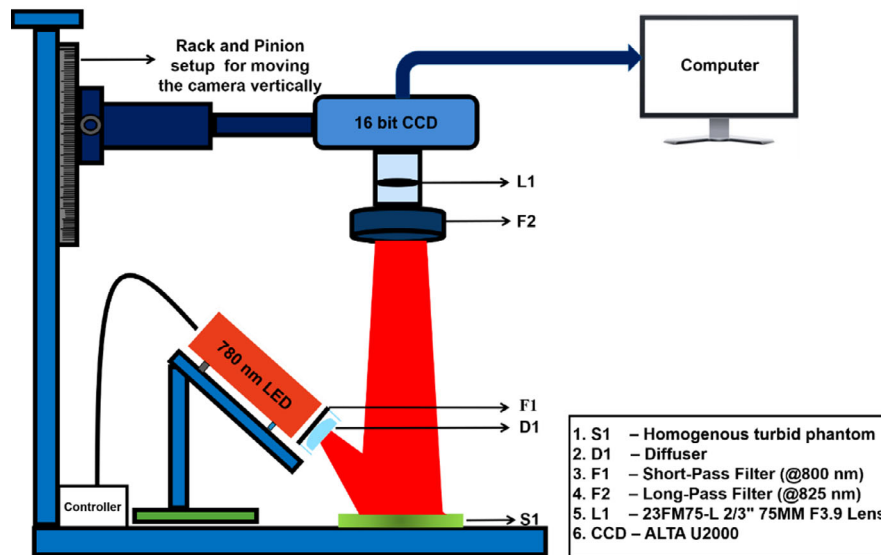


FIG. 1. Schematic of the custom near-infrared fluorescence imaging system.

the phantom were then estimated using the inverse-adding doubling technique.⁴⁵

2.B.1. Multiwell phantom

A multiwell phantom was prepared using thirteen different concentrations of ICG from 0.008 μM to 52 μM . Each sample included ICG, TiO_2 (7.4 mM), resin, and hardener. The protocol for preparing the phantom was similar to that described previously. Mixtures were then cured in separate wells of a 96-well black microplate (Thermo Fisher Scientific, Waltham, MA). This phantom was used for characterizing system sensitivity, linearity, and excitation light crosstalk.

2.B.2. Wide-field phantom

A homogenous ICG-doped turbid phantom was used for characterizing spatial resolution, DOF, signal linearity, FOV, and uniformity. The recipes and experimental protocol for preparing this phantom are similar to that given above. For this wide-field phantom, the ICG concentration was 32.3 μM , the TiO_2 concentration was 152 mM, and the total volume was 30 mL. The prepared final mixture after sonication was poured into 3" \times 6" \times 1.17" molds (Environmental Technology Inc., Fields Landing, CA).

2.B.3. Multichannel phantom

The third type of phantom used in this study leveraged our prior work in the fabrication of 3D-printed phantoms.^{46,47} The phantom was designed in SolidWorks (Dassault Systèmes, Waltham, MA) and printed on a Polyjet printer (Objet260, Stratasys Ltd., Eden Prairie, MN) using proprietary, UV-cured photopolymers. The phantom was printed using white material (VeroWhite, Stratasys Ltd., Eden Prairie, MN) that mimics tissue scattering. Absorption and reduced scattering coefficients at 820 nm were measured to be 0.015 mm^{-1} and 0.52 mm^{-1} , respectively, using a spectrophotometer with an integrating sphere and inverse-adding doubling software.⁴⁵ To avoid signal crosstalk, walls of highly absorbing black material (VeroBlack, Stratasys Ltd., Eden Prairie, MN) were printed between channels (Fig. 2), utilizing the dual-material printing capability of the printer. The diameter of each channel was 2 mm with the total dimension of the phantom being 6.5 \times 3 \times 3 cm. Channels were located at 2, 4, 8, 12, and 16 mm below the surface. ICG (3.2 μM) along with human serum albumin (7.25 μM) was dissolved in PBS and injected into the channels to produce fluorescence contrast.²⁹ This phantom was used to evaluate the penetration depth sensitivity of the NIRF imaging system.



FIG. 2. 3D-printed multichannel phantom for penetration depth measurements. [Color figure can be viewed at wileyonlinelibrary.com]

2.C. Phantom characterization

The fluorescence emission spectra of liquid and solid ICG phantoms were measured with a spectrofluorometer (PTI QuantaMaster QM4, Horiba Scientific, Piscataway, NJ). The ICG concentration for both samples was 3.2 μM . The liquid phantom exhibited an emission with a peak blue shift of approximately 20 nm (from 820 to 800 nm) when it cured. ICG can generate different spectral profiles depending on the solvent used.¹ As ICG belongs to the carbocyanine group, it tends to form aggregates depending upon the concentration and nature of the solvent.⁴⁸

Fluorophore photobleaching can adversely impact standardized testing.²² Therefore, photostability was evaluated to ensure that phantoms produced consistent fluorescence emission over time. The highest concentration in the multiwell phantom (3.2 μM of ICG) was imaged under 740 $\mu\text{W}/\text{cm}^2$, 6 mW/cm^2 , and 12 mW/cm^2 irradiance levels. Results indicated a high degree of stability over a duration relevant to device testing for the two lower intensity levels, whereas a decrease in signal of nearly 10% was seen for the 12 mW/cm^2 case over a period of 30 min.

Long-term change in signals measured from fluorophore-doped polymer phantoms is well documented,^{14,49} and may limit usability over time. In order to assess stability, two types of epoxy-resin phantom were constructed: (a) a multiwell phantom (3.2 μM ICG) and (b) a homogenous turbid phantom (32.3 μM ICG). Fluorescence intensities for both phantoms were recorded with the NIRF imaging system weekly over a period of eight weeks. The multiwell epoxy-resin phantom exhibited $\sim 7\%$ decrease in intensity and the homogenous phantom exhibited $\sim 10\%$ decrease in intensity. Both phantoms exhibited high stability over about one month before fluorescence intensity started to decrease.

2.D. Image quality characteristics

2.D.1. Spatial resolution

The spatial resolution, or sharpness, of an imaging system is a fundamental image quality characteristic that is critical for assessing the ability of a system to resolve fine structures. A variety of approaches for determining spatial resolution are well established for white light imaging systems like endoscopes.⁵⁰ The International Organization for Standardization (ISO) endoscope standard recommends the use of a bar-chart resolution target (e.g., USAF 1951 target) to visually identify resolution in horizontal and vertical directions at the center and four off-axis positions. More objective versions of this approach are suitable for digital imaging systems, such as determination of the contrast transfer function (CTF), which provides data on contrast as a function of spatial frequency.¹⁵ Previous NIRF imaging studies have used versions of the USAF 1951 resolution target that are either diffusely illuminated from behind²² with narrow band light or from the front with white light²¹ or placed atop a homogeneous fluorescent background.²³ Some of these studies have used resolution

targets to identify CTF graphs,²¹ whereas others use them for a more qualitative assessment.²² In a recently published comprehensive multiuse phantom,¹⁶ a resolution target embedded in a turbid matrix was shown to be unusable for determining CTF during surface imaging. Alternate approaches for evaluating NIRF system resolution have included the use of small parallel fluorophores and crossed tubes.²³

In this study, we use a version of the standard bar chart approach in which a negative target (USAF 1951 chrome on glass, Edmund Optics, Barrington, NJ) is placed on top of the wide-field fluorescent phantom (Section 2.D.2), as shown in Fig. 3. The 780 nm LED light source was then used to illuminate the target, and CTFs were calculated using the formula:

$$C_I = \frac{(I_{max} - I_{min})}{(I_{max} + I_{min})} \quad (1)$$

A CTF curve was generated for both horizontal and vertical directions.²¹

2.D.2. Depth of field (DOF)

The DOF of a NIRF imager is important for understanding how image quality is impacted by the camera-target spacing and nonplanar tissue surfaces.^{35,36} DOF was determined by performing spatial resolution measurements (as described in Section 2.D.2) over a range of camera-to-target working distances. CTFs were generated for vertical and horizontal directions. We examined three metrics for DOF, based on: (a) the Rayleigh criterion, (b) a third-order polynomial fit, and (c) contrast at a spatial frequency of 2 lp/mm.

The first method involves determining the spatial frequency at which contrast reaches 26.4% through linear interpolation, at each working distance. For the second approach, a third-order polynomial was used to fit CTF functions and determine resolution. Given that determination of CTF curves can be time-consuming and excessively detailed, a simpler approach was used in which contrast for a single spatial frequency with a moderately high contrast level at best-focus position — 2 lp/mm in this case — is determined as a function of working distance.

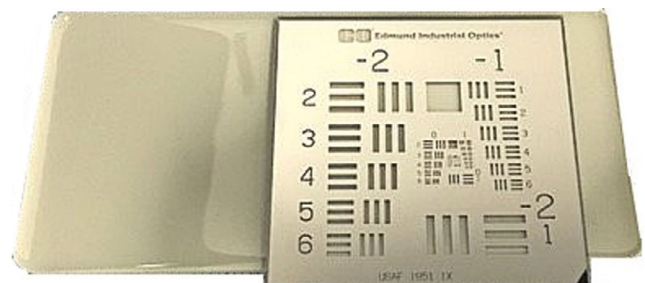


FIG. 3. USAF 1951 resolution test target, negative version, chrome on glass (on top of the wide-field fluorescent phantom). [Color figure can be viewed at wileyonlinelibrary.com]

2.D.3. Sensitivity

The literature includes numerous studies in which samples producing multiple fluorescence intensity levels are used to evaluate performance characteristics such as sensitivity^{1,7,11-16,22,24,26,32} and signal linearity.^{11,13,17,32,33} In most cases, sensitivity of the imaging system is evaluated using multisample phantoms providing a range of fluorophore concentrations.^{14,15} These phantoms should provide biologically relevant turbidity as well as a similar quantum yield level and spectral characteristics (e.g., excitation and emission spectra) as during *in vivo* measurements to optimize clinical relevance of test results. Evaluation of these characteristics is particularly important in polymer phantoms due to the nonbiological environment.

We used an ICG-doped multiwell phantom for this test (described in Section 2.B.1). Each image of the phantom was processed with flat-field correction to correct for uneven illumination of the sample using the following relation:

$$I_3(x, y) = \frac{I_1(x, y)}{I_2(x, y)} \times k \quad (2)$$

where, I_1 = experimental image; I_2 = reference image; k = mean fluorescence intensity of the experimental image;

The mean fluorescence intensity was then calculated over a circular area of 50-pixel diameter centered on the maximum intensity location. The signal to noise ratio (SNR) was calculated by measuring the well with no fluorophore concentration as a mean background value (S_B), mean fluorescence intensity (S_I), and standard deviation of the background well ($\sigma(S_B)$). The SNR was then calculated using the following relation¹¹:

$$\text{SNR} = \frac{S_I - S_B}{\sigma(S_B)} \quad (3)$$

2.D.4. Limits of detection and quantification

One of the most important aspects of sensitivity is the detection limit of a particular fluorophore for a specific imaging system. Detectability has long been a significant issue in analytical chemistry, where concepts such as the limit of detection (LOD) and limit of quantification (LOQ) have been developed to ensure standardized evaluation of system operation.⁵¹ Several other methods have been proposed to explore the minimum sensitivity level of the imaging system, but none has become a standard methodology/protocol.^{1,7,11,12,14,15} Based on the International Standard for Harmonization (ICH) guidelines, LOD is defined as the concentration corresponding to $\text{SNR} = 3$, and LOQ is defined as the concentration corresponding to $\text{SNR} = 10$. We have adhered to these definitions in our analysis. Previously, Gillet *et al.* used this technique in peptide analysis to determine the LOD and LOQ for various systems.⁵² Additionally, it is worth noting that Davis *et al.* used a contrast-to-noise (CNR) of 3 to define the lowest detectable signal of a fluorescence tomography image.³¹

2.D.5. Signal linearity

Signal linearity^{11,13,17,33} is a key characteristic to ensure that the data acquired accurately represents the imaged scene, and that local features of higher intensity (e.g., tumors) can be optimally discriminated from the background. Most commonly, a multiwell phantom approach is used to quantify linearity by fitting the fluorophore concentration vs measured intensity plot. In this paper, linearity is quantified using two test methods: (a) variable fluorophore concentration and (b) variable transmittance. Measurements were performed at multiple exposure durations to assess variations in device performance. The most common^{11,13,17,32,33} approach to quantify linearity involves the use of multiwell phantoms that include fluorophore solutions from a concentration below the LOD to the maximum biologically relevant level.

The second method used to characterize system linearity involves neutral density (ND) filters placed on top of the wide-field phantom, which provides different levels of detected light intensity as with the multiwell approach. An ND filter-based approach was previously used to determine the linearity in a combined Optical Coherence Tomography and Autofluorescence (OCT-AF) imaging system.⁵³ The principle of this approach and the experimental design are shown in Fig. 4. In principle, the emitted fluorescence signal is collected after NIR light is transmitted twice (T^2) through the ND filter. The wide-field phantom was used with a black plastic sheet with a hole in the center (equivalent to the neutral density filter diameter $\sim 0.5''$) and ND filters (OD: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 1.0, 2.0, 3.0 and 4.0; Thorlabs Inc., Newton, NJ) to cover the aperture. Images were then captured with a 1 s exposure for each filter. Furthermore, the transmittance of each filter was calibrated individually using a UV-VIS spectrophotometer (Lambda 1050, Perkin Elmer, Waltham, Massachusetts).

There are two relevant components to evaluation of linearity — the concentration range over which the relationship is linear and the quality of the linear fit over this range. Reporting of both metrics provides the most thorough insight into device performance. The regression starting point should be at the lowest intensity possible, such as the LOD, and include a minimum of five concentration points initially. Subsequent data points are added, and linear regression analysis repeated as long as the R^2 value does not decrease below 0.98. The best R^2 value was reported after comparison between fitted values with respect to different datasets for each exposure time.

2.D.6. Penetration depth

Numerous studies have been published that address how fluorescent inclusion depth impacts intensity and/or apparent inclusion size.^{16,24,26,39} This type of testing is relevant to estimating the thickness of tissue through which a biological structure can be detected, as well as the rate at which depth degrades a reader's ability to assess morphology. For this test, we utilized the 3D-printed, multichannel phantom described

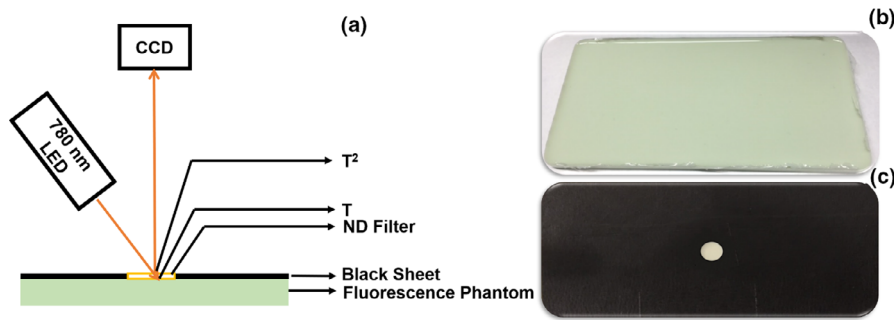


FIG. 4. (a) Schematic of the setup for evaluating sensitivity based on variable transmittance; (b) the wide-field phantom; and (c) wide-field phantom covered with black material. [Color figure can be viewed at wileyonlinelibrary.com]

in Section 2.B.3. Fluorescence images were recorded for each inclusion depth and were used for quantifying three different metrics — the full width at half maximum (FWHM), contrast-to-noise ratio (CNR), and limit of penetration depth. The FWHM analysis was performed by obtaining the horizontal plot profile over an area of 15×482 pixels with the ROI centered on the maximum peak intensity of the fluorescence image. The CNR was quantified as per the reference image protocol outlined (Fig. 5). Subsequently, to find detectable limit of an object, the rose criterion is applied ($CNR > 5$).⁵⁴ The following formula was used to compute the CNR values.

$$CNR = \frac{|S_A - S_B|}{\sigma_0} \tag{4}$$

where S_A = mean signal intensity at the channel; S_B = mean signal intensity in the background; and σ_0 = standard deviation of the background.

2.D.7. Signal uniformity

Evaluation of image uniformity helps to assess the degree to which the device accurately reproduces the spatial distribution of fluorophores. Nonuniform imaging has the potential to limit the useable FOV and alter a reader’s perception of features and trends in an image. Furthermore, it can adversely impact other image quality characteristics such as spatial resolution. Accurate uniformity testing may provide data to perform intensity correction across the image field. Variations in signal intensity across the image field are typically due to the

radially varying intensity of illumination light at the sample surface, although nonideal behavior in the detection path (e.g., vignetting) can contribute significantly as well. Several different methods have been used to characterize signal uniformity,²² some of which collect more sparse data and make assumptions regarding its form (e.g., that it has a Gaussian shape symmetric about the center of the image). Given the potential for irregularities in nonuniformity and the significance of uniformity correction on device performance, a thorough approach that makes no *a priori* assumptions of symmetry and samples the entire FOV is the most appropriate. Toward this end, we have utilized the wide-field phantom described previously. The horizontal and vertical profile was recorded, graphed, and analyzed quantitatively in terms of fractional variation of intensity across the image field.

2.D.8. Excitation light crosstalk

Excitation light leakage into the detection path is a common fluorescence imaging system artifact due to inadequate/deteriorating spectral filtering, and/or the need to detect fluorescence at wavelengths close to the excitation band. A variety of publications have evaluated “crosstalk” effects.^{22,28,38} Zhu *et al.* carried out an experiment with and without the appropriate optics and filters to quantify the leakage level using a phantom.³⁸ In another study, Weiler *et al.* calculated the excitation light leakage under different conditions: (a) shutter closed, (b) phantom with scatterer only, and (c) phantom with a low fluorophore concentration.²⁹ In this study, the following measurements of the multiwell phantom were

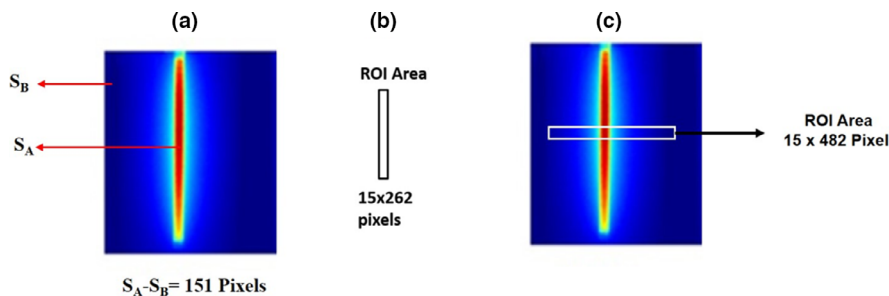


FIG. 5. NIRF images of penetration depth phantom illustrating (a) the protocol for estimating contrast-noise ratio; (b) the dimensions of the sampling area; and (c) the full width at half maximum sampling area.

TABLE I. Overview of phantom-based test methods implemented in this study.

Image quality characteristic	Metric(s)	Phantom	Notes
Sharpness, or spatial resolution	CTF graph, Rayleigh criterion,	Wide-field + USAF 1951 target ^{15,21,29}	Horizontal and vertical directions, ideally at center and near-edge locations
Depth of field (DOF)	CTF, Rayleigh criterion, 2 lp/mm contrast vs depth	Wide-field + USAF 1951 target ³⁵⁻³⁷	May require disabling of autofocus routines
Sensitivity	Graph signal vs concentration, LOD, LOQ	Multiwell phantom ^{7,11,15-17,22,24,27} (e.g., 12-well)	Fluorophore properties may be environment-dependent
Linearity	Concentration range, R ² Graph of signal vs transmittance, range, R ²	Multiwell phantom ^{11,13,17,32,33} (e.g., 12-well) Wide-field + ND filters ⁵³	Potential concentration-dependent nonlinearities Less dependent on sample, but lacks direct correlation to <i>in vivo</i> concentrations
Penetration depth	Intensity, FWHM vs depth, Depth for CNR > 5 ⁵⁴	Multichannel phantom ²⁹	Highly dependent on concentration, but method avoids errors due to cross-talk
Field of view (FOV)	Dimensions (cm)	Wide-field phantom or grid	Measure directly from image
Signal uniformity	Graph, % variation	Wide-field ^{16,24,34}	Includes illumination and imaging nonuniformity
Excitation light crosstalk	Fluorescence intensity (counts)	Multiwell phantom (4-well) ^{16,24,25,29,38,39}	Depends on illumination and collection optics

performed: (a) closed shutter, (b) wells with a high scattering ($\mu_s' \sim 20 \text{ cm}^{-1}$ at 800 nm) but no ICG, and (c) wells with biologically relevant ICG concentration.

A summary of the metrics, phantoms, and relevant literature for each of the described image quality characteristics is provided in Table I.

3. RESULTS AND DISCUSSION

3.A. Spatial resolution

The CTF curves for best focus were plotted for horizontal and vertical directions and fitted with third-order polynomials (Fig. 6) that provided high-quality representations ($R^2 = 0.99$). The Rayleigh criterion was then applied to determine the spatial resolution of our imaging system.^{15,21} Horizontal and vertical CTF curves appeared very similar, yet the respective resolution values were slightly different, at 0.31 mm (3.2 lp/mm) and 0.29 mm (3.5 lp/mm).

While our test method is less compact and expedient than the L-shaped fluorescence strips used in a prior “all-in-one” performance testing phantom,¹⁶ it is a more well established

method with results that are both readily achieved with commercial targets and widely understood. For these reasons, our approach is more well suited to occasional, rigorous performance characterization, whereas a compact multiple-characteristic phantom may be more practical for day-to-day testing. One potential challenge with our approach is the need to ensure that the wide field fluorescent phantom is flat, homogeneous and maintains direct contact with the USAF 1951 target during measurements.

3.B. Depth of field (DOF)

Measurements of DOF provide insight into the sensitivity of image resolution to variations in device-target distance.³⁵⁻³⁷ Results in Fig. 7 show that as the phantom is moved toward the position of best focus from a location 18 mm away, contrast across the entire range of spatial frequencies increases gradually, then decreases again once the phantom passes through the focal plane. This result indicates that imaging of nonplanar surfaces that vary on the order of 1 cm relative to the focal plane may be impacted by reductions in resolution. Additionally, Fig. 7 shows that the CTF curves for

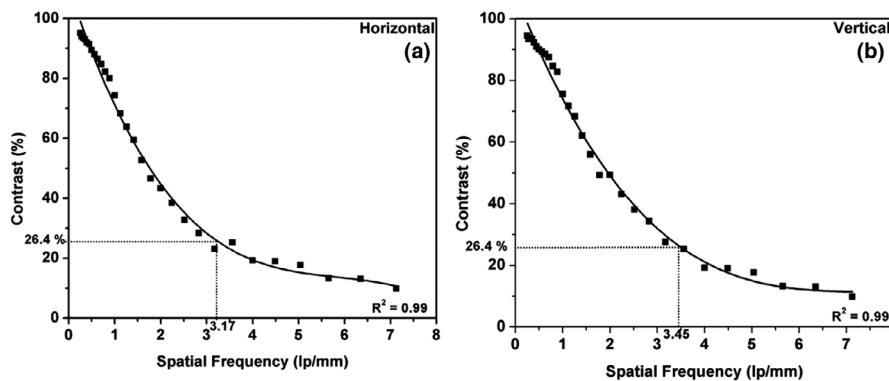


FIG. 6. Spatial resolution results presented in the form of contrast transfer functions for (a) horizontal and (b) vertical directions.

corresponding locations on either side of the focal plane are not identical, and that curves for positions closer to the focal plane show higher contrast at higher spatial frequencies than for positions further from the focal plane.

Figure 8 shows results for Rayleigh resolution values estimated directly from measured results as well as values determined from CTF curves fitted with a third-order polynomial [see Fig. 8(b)]. The alternate, simpler approach is also presented in which contrast is provided for a single set of bars at

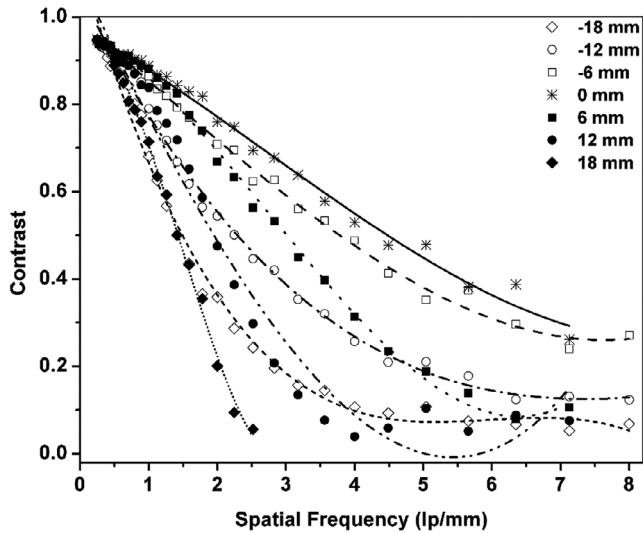


FIG. 7. Depth of field results showing contrast transfer functions at seven locations (vertical direction).

a spatial frequency of 2.0 lp/mm as the camera-to-target distance is varied. While the single-frequency resolution target approach is simpler, it does not show as strong a distinction between in-focus and out-of-focus regions. Among the techniques evaluated, the Rayleigh criterion-based approaches are the more repeatable and reliable methods to explore DOF.

3.C. Sensitivity

Individually normalized images for each ICG concentration (0.008–52 μM) using three exposure times (100, 500, and 1000 ms) are shown in Fig. 9. Normalizing the fluorescence images helps to illustrate the difference in fluorescence intensity caused by fluorophore concentration. For example, fluorescence images taken of 13 and 0.13 μM ICG concentration phantoms at 100 ms show distinct differences. On the other hand, fluorescence images taken at a higher concentration range (26–52 μM) seem to be similar. Hence, the mean fluorescence intensity and signal-to-noise ratio against concentration (Fig. 10) were quantified from Fig. 9.

As illustrated by Fig. 10(a), mean fluorescence intensity increases proportionately with ICG concentration over most of the phantom's range. For example, fluorescence intensity for 1000 ms exposures shows a linear trend until 26 μM , after which the slope of the curve decreases. Similarly, a linear relationship is seen until 52 μM for 100 and 500 ms exposures. Figure 10(b) shows the corresponding SNR as a function of concentration. The SNR is used further to quantify the LOD and LOQ.

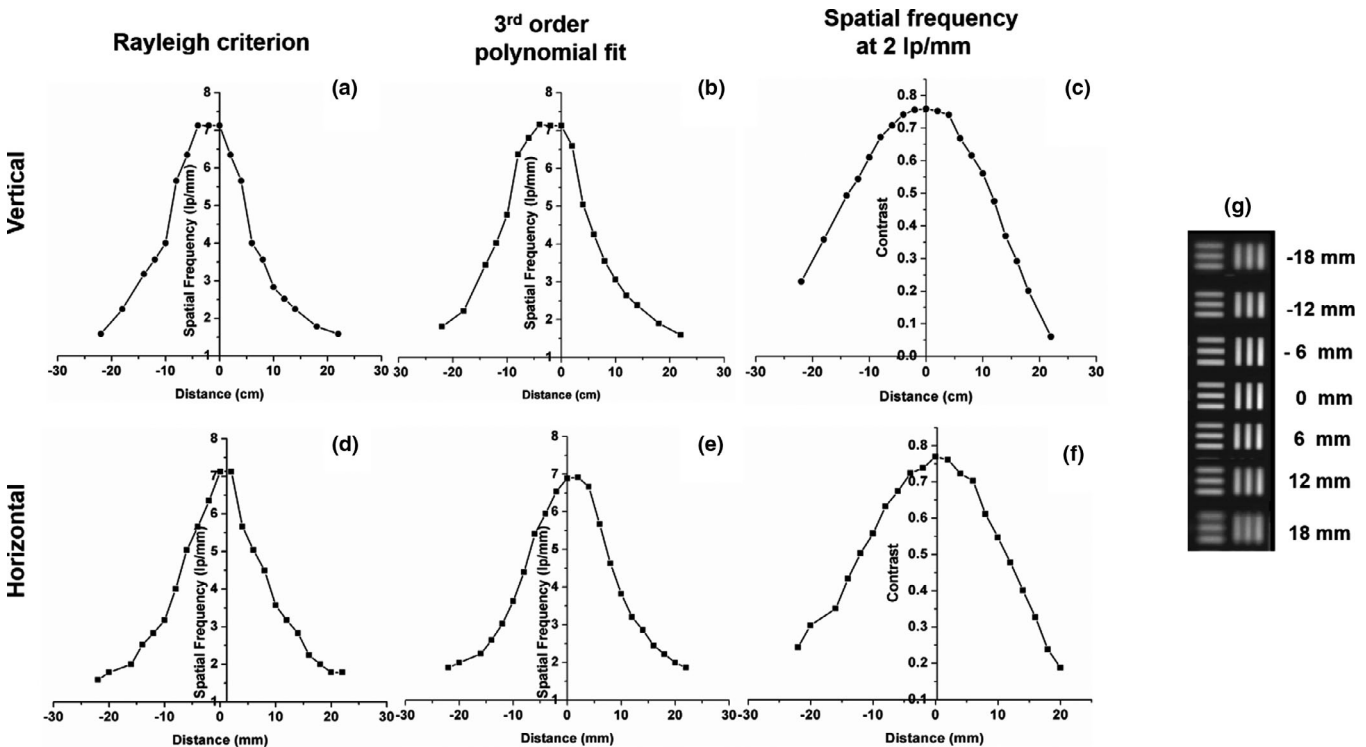


FIG. 8. Results for depth of field based on different methods: (a) Rayleigh criterion (without fit) (b) third-order polynomial fit and (c) contrast for a spatial frequency of 2 lp/mm in vertical direction, as well as corresponding results for the horizontal direction (d)–(f). An example set of images for 2 lp/mm is shown in (g).

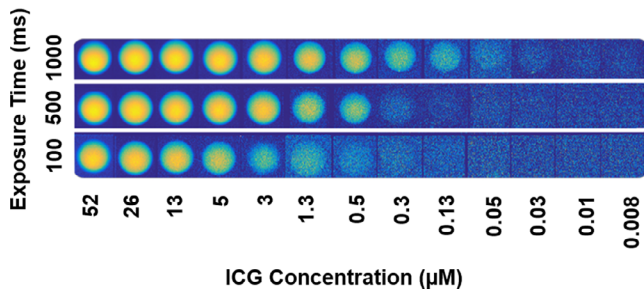


FIG. 9. Individually peak-normalized images of indocyanine green-doped epoxy-resin wells, acquired at three exposure durations (100, 500, and 1000 ms).

It is well known that the fluorescence characteristics of ICG depend upon both the concentration and solvent used. Weiler *et al.*²⁹ reported that the fluorescence intensity of ICG in albumin increased with dye concentration, reaching a maximum at 258 μM before decreasing. In this study, an ICG concentration of 52 μM was determined to be the limit of the linear range. In one prior phantom-based sensitivity study quantum dots were used to provide a widely useful surrogate fluorophore.¹¹ While such a fluorophore may generate significant signal across a large spectral range, differences in the excitation and emission properties between quantum dots and ICG — or any other fluorescence dye to be imaged clinically by a device — may cause the test to yield unrealistic device performance comparisons. Obtaining sufficient concentrations of quantum dots to enable testing across the relevant sensitivity range may also be prohibitively expensive. Additionally, it is worth noting that sensitivity test methods such as these involving variations in contrast agent concentration can be applied in an analogous manner to other optical modalities, including emerging approaches like photoacoustic imaging^{55–58} where the use of subsurface inclusions with varying chromophore concentrations has been employed.

3.C.1. Limit of detection and limit of quantification

Prior studies have proposed nonstandardized protocols for determining sensitivity limits of a fluorophore from multiwell measurements. As described earlier, our protocol for

determining limit of detection and quantification is based on ICH guidelines.⁵¹ Descriptions for the quantification of LOD and LOQ have been outlined previously (Section 2.D.4). Based on these, we have determined results for different exposure durations, shown in Table II.

3.D. Signal linearity

3.D.1. Variable fluorophore concentration

Imaging system linearity plays a vital role in determining the degree to which fluorescent structures are visible to the user. In this study, linear regression analysis of multiwell phantom measurement results was performed, and the linearity of the imaging system for exposure times of 100, 500 and 1000 ms are reported. The variance of linearity with respect to concentration was reported as the R^2 value of the linear fit. The ICG concentration range from 0.5 to 13 μM exhibited $R^2 \sim 0.98$ for all exposure times. Hence, the 0.5–13 μM concentration range can be used for quantification using the NIRF imaging system. The calibration curves for all three exposure times showed good linearity ($R^2 = 0.98$ – 0.99), provided the upper concentration limit was set to 13 μM . A similar technique has been used for fluorescence tomography as well.⁵⁹ Alternative data analysis methods have also been reported in the literature, such as determining the slope of a linear fit to a log-log plot,³⁵ yet no standardized thresholds have been established for NIRF imaging.

3.D.2. Variable transmittance approach

As described previously (Section 2.D.5), the second sensitivity approach does not depend on changes in fluorophore concentration. A compilation of images of individual wells, using consistent normalization [Fig. 11(a)] and image-specific normalization [Fig. 11(b)] show the decrease in NIRF signal intensity as ND filter attenuation increases [Figs. 11(a)–11(b)]. A log-log plot of mean intensity with respect to Transmittance² is shown in Fig. 11(c). As the OD level reaches 4 and 5 (approaching noise level), the variation in mean fluorescence intensity is similar [Fig. 11(c)]. Regression analysis indicated a high degree of linearity ($R^2 = 0.98$) once signals

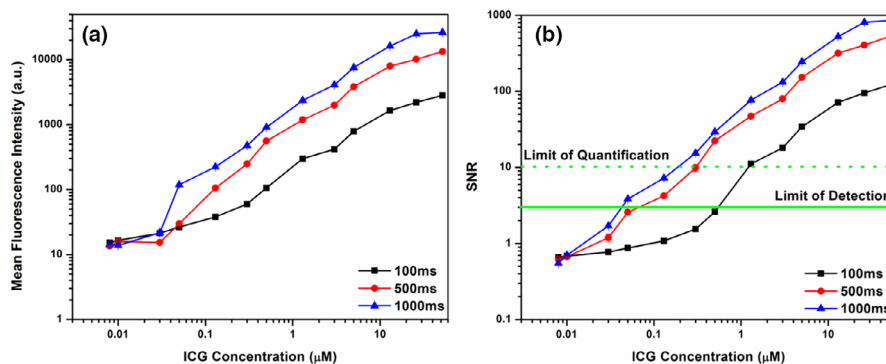


FIG. 10. Results for mean fluorescence intensity (a) and signal-to-noise ratio (b) in an ICG-doped, epoxy-resin multiwell phantom. Mean measurements ($n = 3$) are shown, yet the standard deviation was not significant enough to show error bars.

TABLE II. Results for LOD and LOQ in ICG-doped sensitivity phantoms.

Parameter (μM)	Exposure time (sec)		
	0.1	0.5	1
Limit of detection	0.52	0.12	0.08
Limit of quantification	1.68	0.39	0.26

increased above the noise floor. This type of ND filter approach has been used to evaluate the nonlinearity response of CCD detectors⁶⁰ as well as a combined OCT-autofluorescence imaging system.⁵³ Overall, our results indicate that this is a useful approach to evaluate NIRF imaging system linearity, although it does not provide direct insight into fluorophore nonlinearity or detection limit. In many cases, the ND filter and multiwell phantom approaches may provide useful complementary insights into device performance.

3.E. Penetration depth

In order to assess the impact of inclusion depth on detected images, we have generated data on two metrics derived from the same images of fluorophore-filled channels at multiple depths [Fig. 12(a)] — signal intensity and inclusion width. While fluorescence intensity distributions of channels at depths of 2 and 4 mm appear similar, results for depths of 8 mm or more exhibit increasingly broad channels, and greater intensity in nonfluorescent regions. Quantitative data derived from these images for signal intensity and channel width are shown in Figs. 12(b) and 12(c), respectively. A fit analysis of the channel width plot indicates strong linearity with depth ($R^2 = 0.99$), which is in agreement with prior studies.^{26,61,62}

As illustrated in Fig. 13, CNR decreases rapidly with channel depth until 8 mm (17.7 ± 0.3). After 8 mm, the fluorescence image appears blurred, yet it was possible to differentiate the general region of the channel [Fig. 12(a)]. The CNR values for 12 and 16 mm depth channels were 5.2 ± 0.1 and 2.24 ± 0.03 , respectively. According to the Rose criterion⁵⁴ the minimum CNR value required to detect an object is 3–5. Hence, the limit of penetration depth for the channel phantom is 12 mm. A similar threshold limit (CNR = 3) has also been used to analyze OCT images.³¹ Graphs of SNR and CNR provided similar insights, yet CNR appears more well suited to characterizing imaging performance in this type of test. Overall, the depth-varying inclusion approach is a relatively simple, yet effective method that can be adapted for use in a range of imaging modalities, such as using chromophore-filled inclusions for photoacoustics.^{55–58} The advantage provided by our 3D-printed phantom design incorporating highly absorbing barriers between channels for NIRF imaging is that we are able to essentially eliminate crosstalk between channel regions.

3.F. Signal uniformity

In many applications, uniformity of the light source is necessary to make sure that the light is uniform within the sample illumination area. The 780 nm LED was equipped with the appropriate diffuser to make the incident light uniform (Fig. 14). By acquiring images at multiple positions across the wide-field phantom, we were able to determine that the phantom itself was highly uniform, as lateral changes in signal intensity were not significantly impacted by position. The change in signal uniformity was primarily radial in nature, decreasing from the center (800, 600 pixels) to the edges of the image. A change of 43% was seen from the middle to the

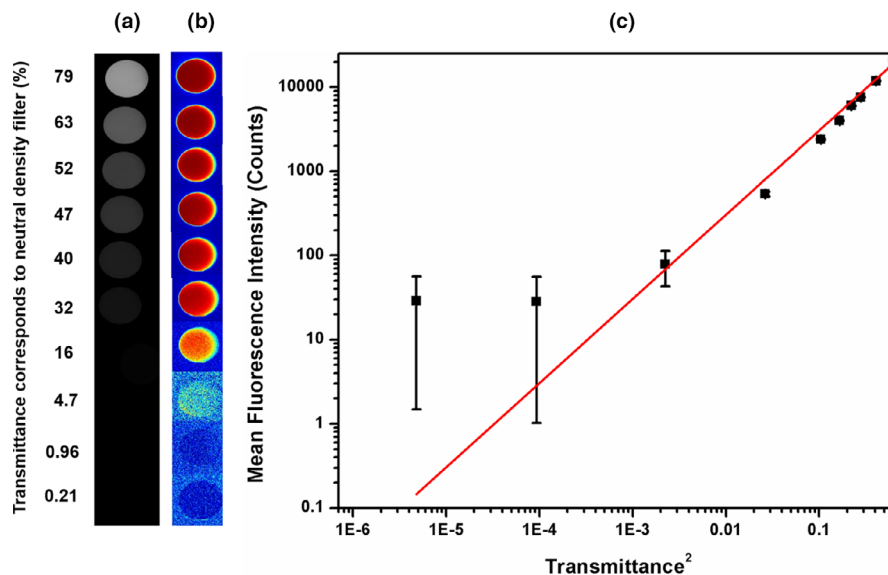


FIG. 11. (a) Gray scale near-infrared fluorescence images as a function of OD value, (b) individually normalized false-color images, and (c) a graphical analysis of results along with linear regression fit.

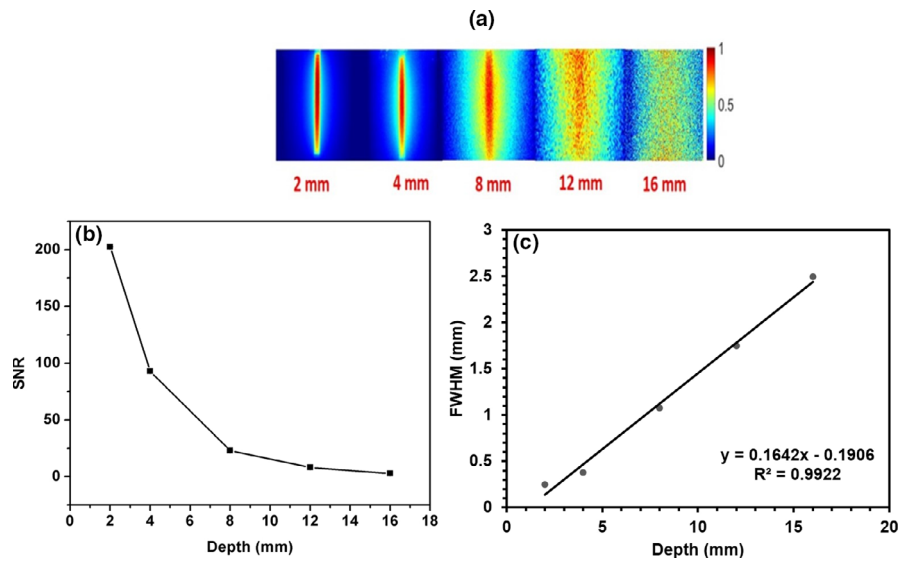


FIG. 12. (a) Individually normalized images of the multichannel phantom, and quantitative analyses of the channel phantom, including (b) signal-to-noise ratio and (c) apparent channel width as a function of depth.

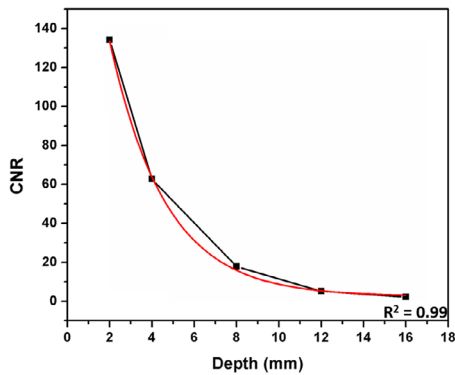


FIG. 13. Contrast-to-noise as a function of depth for the multichannel phantom. [Color figure can be viewed at wileyonlinelibrary.com]

top/bottom, and a change of 51% was seen from the middle to the right/left edges (as defined by the rectangular regions in Fig. 14).

3.G. Excitation light crosstalk

Results are shown in Fig. 15 for the three conditions (Section 2.D.8) used to evaluate the presence of excitation light in fluorescence images. The closed shutter image yielded a mean detector noise level of about 1850 ± 1.9 counts, whereas the nonfluorescent turbid sample showed a mean of 1934 ± 0.3 counts and the fluorescent sample generated a signal level of 4527 ± 2.3 counts. These data indicate that while a small amount of excitation light leakage is present, this level is minor compared to the signal measured at a biologically relevant ICG concentration. Specifically, the crosstalk level was 84 counts above the background noise, whereas the biologically relevant level was a factor of 31 higher, after background subtraction.

4. CONCLUSIONS

Based on a wide range of approaches described in medical and fluorescence imaging literature, we have developed and demonstrated a cohesive battery of test methods for evaluation of fluorescence image quality in wide-field imagers. The following performance characteristics were addressed: spatial resolution, DOF, sensitivity, LOD, LOQ, linearity, penetration depth, FOV, uniformity, contrast-detail analysis, and excitation light leakage. We also propose a number of key metrics that can facilitate direct, quantitative comparison of device performance. Furthermore, several of the above test methods may be modified slightly (e.g., through the use of chromophores instead of fluorophores) for use in emerging modalities such as photoacoustic tomography/microscopy and spatial frequency domain imaging. These methods have the potential to facilitate more uniform evaluation and inter-comparison of clinical and preclinical imaging systems than is typically achieved, with the long-term goal of establishing international standards for fluorescence image quality assessment.

ACKNOWLEDGMENTS

This work was funded by NSF/FDA Scholar-in-Residence Program (CBET-1445701 and CBET-1641077). The authors gratefully acknowledge FDA colleagues, including Dr. Kejing Chen for input on performance testing, and Dr. Jianting Wang and Dr. William C. Vogt for advice on phantom preparation. Three-dimensional printing was performed with the assistance of Mr. Gonzalo Mendoza in the FDA Additive Manufacturing of Medical Products (AMMP) Lab, supported by the Office of the Chief Scientist's Shared Resources program.

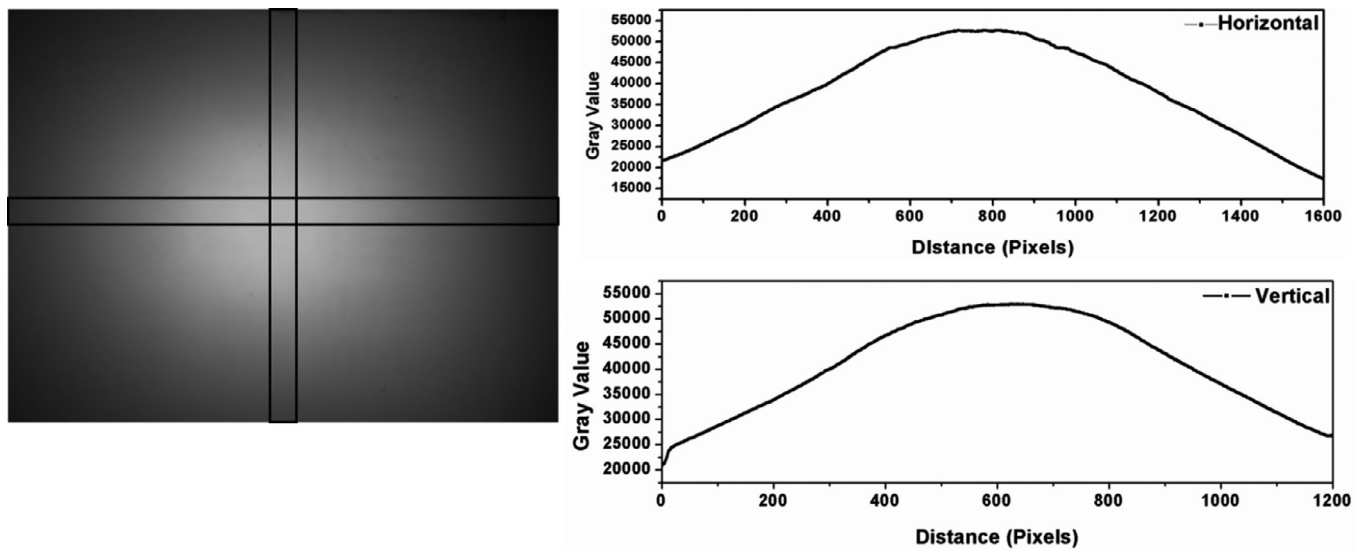


FIG. 14. Uniformity is illustrated in terms of two-dimensional spatial distribution as well as horizontal and vertical profiles.

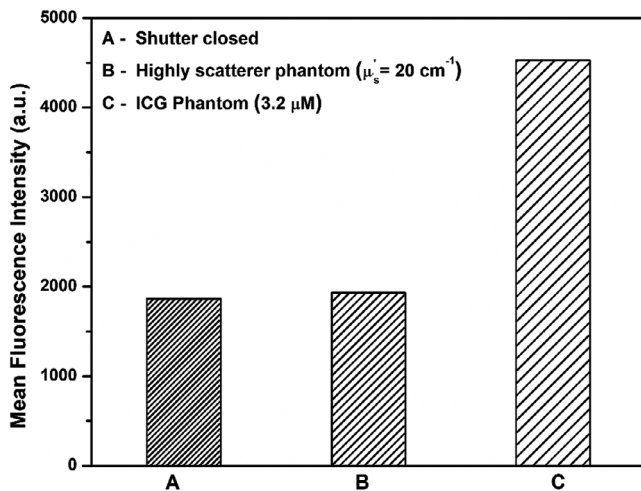


FIG. 15. Excitation light leakage measurements including results for the following cases: (a) shutter closed, (b) nonfluorescent epoxy-resin phantom with $\mu_s' = 20 \text{ cm}^{-1}$ and (c) epoxy-resin phantom with biologically relevant indocyanine green concentration ($3.2 \text{ } \mu\text{M}$).

5. DISCLAIMER

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services, and National Institute of Standards and Technology.

CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

^{a)}Authors to whom correspondence should be addressed. Electronic mails: Joshua.Pfefer@fda.hhs.gov, ychen8@umass.edu.

REFERENCES

- Zhu B, Sevick-Muraca EM. A review of performance of near-infrared fluorescence imaging devices used in clinical studies. *Br J Radiol.* 2015;88:20140547.
- Sevick-Muraca EM, Sharma R, Rasmussen JC, et al. Imaging of lymph flow in breast cancer patients after microdose administration of a near-infrared fluorophore: feasibility study. *Radiology.* 2008;246:734–741.
- Yuan J, Yi X, Yan F, et al. Near-infrared fluorescence imaging of prostate cancer using heptamethine carbocyanine dyes. *Mol Med Rep.* 2015;11:821–828.
- Unno N, Nishiyama M, Suzuki M, et al. quantitative lymph imaging for assessment of lymph function using indocyanine green fluorescence lymphography. *Eur J Vasc Endovasc Surg.* 2008;36:230–236.
- Eisenberg DP, Adusumilli PS, Hendershott KJ, et al. Real-time intraoperative detection of breast cancer axillary lymph node metastases using a green fluorescent protein-expressing herpes virus. *Ann Surg.* 2006;243:824–830.
- Tanaka E, Choi HS, Fujii H, Bawendi MG, Frangioni JV. Image-guided oncologic surgery using invisible light: completed pre-clinical development for sentinel lymph node mapping. *Ann Surg Oncol.* 2006;13:1671–1681.
- Meric-Bernstam F, Rasmussen JC, Krishnamurthy S, et al. Toward nodal staging of axillary lymph node basins through intradermal administration of fluorescent imaging agents. *Biomed Opt Express.* 2014;5:183.
- Ihalainen TM, Lönnroth NT, Peltonen JI, et al. MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncol.* 2011;50:966–972.
- Weisser G, Lehmann KJ, Scheck R, Coppenrath E, Georgi M. Dose and image quality of electron-beam CT compared with spiral CT. *Invest Radiol.* 1999;34:415–420.
- Bouchet F, Geworski L, Knoop BO, et al. Calibration test of PET scanners in a multi-centre clinical trial on breast cancer therapy monitoring using 18F-FLT. *PLoS ONE.* 2013;8:e58152.
- Zhu B, Rasmussen JC, Sevick-Muraca EM. A matter of collection and detection for intraoperative and noninvasive near-infrared fluorescence molecular imaging: to see or not to see? *Med Phys.* 2014;41:1–11.
- Li X, Li Z, Gan W, et al. Sensitive and high resolution subcutaneous fluorescence in vivo imaging using upconversion nanoparticles and microarrays. *Analyst.* 2013;138:3711–3718.
- Gan Q, Wang D, Ye J, et al. Benchtop and animal validation of a projective imaging system for potential use in intraoperative surgical guidance. *PLoS ONE.* 2016;11:1–21.
- Zhu B, Tan I-C, Rasmussen JC, Sevick-Muraca EM. Validating the sensitivity and performance of near-infrared fluorescence imaging and tomography devices using a novel solid phantom and measurement approach. *Technol Cancer Res Treat.* 2012;11:95–104.

15. Zhu B, Rasmussen JC, Litorja M, Sevick-Muraca EM. Determining the performance of fluorescence molecular imaging devices using traceable working standards with SI units of radiance. *IEEE Trans Med Imaging*. 2016;35:802–811.
16. Anastasopoulou M, Koch M, Gorpas D, et al. Comprehensive phantom for interventional fluorescence molecular imaging. *J Biomed Opt*. 2016;21:91309.
17. Ghassemi P, Wang B, Wang J, Wang Q, Chen Y, Pfefer TJ. Evaluation of mobile phone performance for near-infrared fluorescence imaging. *IEEE Trans Biomed Eng*. 2017;64:1650–1653.
18. Wang H, Feng X, Shi B, et al. Signal-to-noise ratio analysis and improvement for fluorescence tomography imaging. *Rev Sci Instrum*. 2018;89:093114.
19. Ghassemi P, Wang J, Melchiorri AJ, et al. Rapid prototyping of biomimetic vascular phantoms for hyperspectral reflectance imaging. *J Biomed Opt*. 2015;20:121312.
20. Brown EN, Burris NS, Gu J, et al. Thinking inside the graft: applications of optical coherence tomography in coronary artery bypass grafting. *J Biomed Opt*. 2007;12:051704.
21. Themelis G, Yoo JS, Soh K-S, Schulz R, Ntziachristos V. Real-time intraoperative fluorescence imaging system using light-absorption correction. *J Biomed Opt*. 2009;14:064012.
22. Kittle DS, Mamelak MDA, Parrish-Novak JE, et al. Fluorescence-guided tumor visualization using the tumor paint BLZ-100. *Cureus*. 2014;6:e210.
23. Graves EE, Ripoll J, Weissleder R, Ntziachristos V. A submillimeter resolution fluorescence molecular imaging system for small animal imaging. *Med Phys*. 2003;30:901–911.
24. Gorpas D, Koch M, Anastasopoulou M, Klemm U, Ntziachristos V. Benchmarking of fluorescence cameras through the use of a composite phantom. *J Biomed Opt*. 2017;22:16009.
25. Chen Z, Zhu N, Pacheco S, Wang X, Liang R. Single camera imaging system for color and near-infrared fluorescence image guided surgery. *Biomed Opt Express*. 2014;5:2791–2797.
26. Pleijhuis RG, Langhout GC, Helfrich W, et al. Near-infrared fluorescence (NIRF) imaging in breast-conserving surgery: assessing intraoperative techniques in tissue-simulating breast phantoms. *Eur J Surg Oncol*. 2011;37:32–39.
27. van Willigen DM, van den Berg NS, Buckle T, et al. Multispectral fluorescence guided surgery; a feasibility study in a phantom using a clinical-grade laparoscopic camera system. *Am J Nucl Med Mol Imaging*. 2017;7:138–147.
28. Volpi D, Tullis IDC, Laios A, et al. A novel multiwavelength fluorescence image-guided surgery imaging system. *Proc SPIE*. 2014:89350G.
29. Weiler M, Kassir T, Dixon JB. Sensitivity analysis of near-infrared functional lymphatic imaging. *J Biomed Opt*. 2012;17:066019.
30. Li X, Chance B, Yodh AG. Fluorescent heterogeneities in turbid media: limits for detection, characterization, and comparison with absorption. *Appl Opt*. 1998;37:6833.
31. Davis SC, Pogue BW, Dehghani H, Paulsen KD. Contrast-detail analysis characterizing diffuse optical fluorescence tomography image reconstruction. *J Biomed Opt*. 2005;10:050501.
32. Gao S, Mondal SB, Zhu N, Liang R, Achilefu S, Gruev V. Image overlay solution based on threshold detection for a compact near infrared fluorescence goggle system. *J Biomed Opt*. 2015;20.
33. Elliott JT, Dsouza AV, Marra K, Pogue BW, Roberts DW, Paulsen KD. Microdose fluorescence imaging of ABY-029 on an operating microscope adapted by custom illumination and imaging modules. *Biomed Opt Express*. 2016;7:3280–3288.
34. Katrašnik J, Pernuš F, Likar B. A method for characterizing illumination systems for hyperspectral imaging. *Opt Express*. 2013;21:4841–4853.
35. DSouza AV, Lin H, Henderson ER, Samkoe KS, Pogue BW. Review of fluorescence guided surgery systems: identification of key performance capabilities beyond indocyanine green imaging. *J Biomed Opt*. 2016;21:080901.
36. Sellors JW, Winkler JL, Kreysar DF. Illumination, optics, and clinical performance of a hand-held magnified visual inspection device (AviScope(TM)): a comparison with colposcopy. *J Acquir Immune Defic Syndr*. 2004;37:S160–S166.
37. Ortyń WE, Perry DJ, Venkatachalam V, et al. Extended depth of field imaging for high speed cell analysis. *Cytom Part A*. 2007;71:215–231.
38. Zhu B, Rasmussen JC, Lu Y, Sevick-Muraca EM. Reduction of excitation light leakage to improve near-infrared fluorescence imaging for tissue surface and deep tissue imaging. *Med Phys*. 2010;37:5961–5970.
39. Houston JP, Thompson AB, Gurfinkel M, Sevick-Muraca EM. Sensitivity and depth penetration of continuous wave versus frequency-domain photon migration near-infrared fluorescence contrast-enhanced imaging. *Photochem Photobiol*. 2003;77:420–430.
40. Wang Q, Chen WC, Suresh N, Hua H. Development of the local magnification method for quantitative evaluation of endoscope geometric distortion. *J Biomed Opt*. 2016;21:056003.
41. Photography — Archiving systems — Image quality analysis — Part 1: reflective originals. ISO/TS 19264-1:2017; Geneva, Switzerland.
42. Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using µManager software. *J Biol Methods*. 2014;1:10.
43. Eliceiri K, Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671–675.
44. Scott C, Potsaid B, Wen JT. Wide field scanning telescope using mems deformable mirrors. *Int J Optomechatronics*. 2010;4:285–305.
45. Prah SA, van Gemert MJC, Welch AJ. Determining the optical properties of turbid media by using the adding-doubling method. *Appl Opt*. 1993;32:559–568.
46. Liu Y, Ghassemi P, Depkon A, et al. Biomimetic 3D-printed neurovascular phantoms for near-infrared fluorescence imaging. *Biomed Opt Express*. 2018;9:2810–2824.
47. Wang J, Coburn J, Liang C-P, et al. Three-dimensional printing of tissue phantoms for biophotonic imaging. *Opt Lett*. 2014;39:3010–3013.
48. Desmettre T, Devoisselle JM, Mordon S. Fluorescence properties and metabolic features of indocyanine green (ICG) as related to angiography. *Surv Ophthalmol*. 2000;45:15–27.
49. Baeten J, Niedere M, Dunham J, Ntziachristos V. Development of fluorescent materials for diffuse fluorescence tomography standards and phantoms. *Opt Express*. 2007;15:8681–8694.
50. Optics and photonics — Medical endoscopes and endotherapy devices — Part 5: determination of optical resolution of rigid endoscopes with optics. ISO 8600-5:2005; Geneva, Switzerland.
51. Validation of analytical procedures: text and methodology Q2(R1). ICH; 2005.
52. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the ms/ms spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics*. 2012;11:016717.
53. Nafar Z, Jiang M, Wen R, Jiao S. Visible-light optical coherence tomography-based multimodal retinal imaging for improvement of fluorescent intensity quantification. *Biomed Opt Express*. 2016;7:3220–3229.
54. Cherry SR, Sorenson JA, Phelps ME. Tomographic reconstruction in nuclear medicine. In: *Physics in Nuclear Medicine*. Philadelphia, PA: Elsevier; 2012.
55. Vogt WC, Zhou X, Andriani R, Wear KA, Pfefer TJ, Garra BS. Photoacoustic oximetry imaging performance evaluation using dynamic blood flow phantoms with tunable oxygen saturation. *Biomed Opt Express*. 2019;10:449.
56. Vogt WC, Wear KA, Garra BS, et al. Phantom-based image quality test methods for photoacoustic imaging systems. *J Biomed Opt*. 2017;22:1.
57. Li M, Liu C, Gong X, et al. Linear array-based real-time photoacoustic imaging system with a compact coaxial excitation handheld probe for noninvasive sentinel lymph node mapping. *Biomed Opt Express*. 2018;9:1408–1422.
58. Bohndiek SE, Bodapati S, Van De Sompel D, Kothapalli S-R, Gambhir SS. Development and application of stable phantoms for the evaluation of photoacoustic imaging instruments. *PLoS ONE*. 2013;8:e75533.
59. Kepshire D, Davis SC, Dehghani H, Paulsen KD, Pogue BW. Fluorescence tomography characterization for sub-surface imaging with protoporphyrin IX. *Opt Express*. 2008;12:8581–8593.
60. Williams TC, Shaddix CR. Simultaneous correction of flat field and nonlinearity response of intensified charge-coupled devices. *Rev Sci Instrum*. 2007;78:123702.
61. Starosolski Z, Bhavane R, Ghaghada KB, Vasudevan SA, Kaay A, Annappagada A. Indocyanine green fluorescence in second near-infrared (NIR-II) window. *PLoS ONE*. 2017;12:1–14.
62. Welscher K, Sherlock SP, Dai H. Deep-tissue anatomical imaging of mice using carbon nanotube fluorophores in the second near-infrared window. *Proc Natl Acad Sci*. 2011;108:8943–8948.