Multi-Parametric Standardization of Fluorescence Imaging Systems Based on a Composite Phantom

Dimitris Gorpas, *Member, IEEE*, Maximilian Koch, Maria Anastasopoulou, Dmitry Bozhko,
Uwe Klemm, Markus Nieberler, and Vasilis Ntziachristos, *Senior Member, IEEE*

Abstract*—*Objective: **Fluorescence molecular imaging (FMI) has emerged as a promising tool for surgical guidance in oncology, with one of the few remaining challenges being the ability to offer quality control and data referencing. This paper investigates the use of a novel composite phantom to correct and benchmark FMI systems. *Methods:* This paper extends on previous work by describing a phantom design that can provide a more complete assessment of FMI systems through quantification of dynamic range and determination of spatial illumination patterns for both reflectance and fluorescence imaging. Various performance metrics are combined into a robust and descriptive "system benchmarking score," enabling not only the comprehensive comparison of different systems, but also for the first time, correction of the acquired data. *Results:* We show that systems developed for targeted fluorescence imaging can achieve benchmarking scores of up to 70%, while clinically available systems optimized for indocyanine green are limited to 50%, mostly due to greater leakage of ambient and excitation illumination and lower resolution. The image uniformity can also be approximated and employed for image flat-fielding, an important milestone toward data referencing. In addition, we demonstrate composite phantom use in assessing the performance of a surgical microscope and of a raster-scan imaging system. *Conclusion:* Our results suggest that the new phantom has the potential to support high-fidelity FMI through benchmarking and image correction. *Significance:* Standardization of the FMI is a necessary process for establishing good imaging practices in clinical environments and for enabling high-fidelity imaging across patients and multi-center imaging studies.**

***Index Terms*—Fluorescence molecular imaging, composite phantom, standardization, benchmarking, data referencing.**

Manuscript received February 7, 2019; revised April 4, 2019; accepted April 5, 2019. Date of publication April 11, 2019; date of current version December 23, 2019. This work was supported in part by the Deutsche Forschungsgemeinschaft, Germany [Gottfried Wilhelm Leibniz Prize 2013; NT 3/10-1] and in part by the Bundesministerium für Bildung und Forschung (BMBF), Bonn, Germany (Project ESCEND, 01KT1809), under the framework of the ERA-NET TRANSCAN-2 initiative under Grant 643638. *(Corresponding author: Vasilis Ntziachristos.)*

D. Gorpas, M. Koch, M. Anastasopoulou, D. Bozhko, and U. Klemm are with the Institute of Biological and Medical Imaging, Helmholtz Zentrum München and also with the Chair of Biological Imaging and TranslaTUM, Technische Universität München.

M. Nieberler is with the Department of Oral and Maxillofacial Surgery, University Hospital Rechts der Isar, Technische Universität München.

V. Ntziachristos is with the Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg 85764, Germany, and also with the Chair of Biological Imaging and TranslaTUM, Technische Universität München, Munich 81675, Germany (e-mail: v.ntziachristos@tum.de).

Digital Object Identifier 10.1109/TBME.2019.2910733

I. Introduction

Fluorescence-guided intervention is increasingly considered for real-time intra-operative oncological applications, such as earlier cancer detection [1] or improving surgical outcome by more accurate delineation of tumor margins [2]–[4]. Indocyanine green (ICG) is the most commonly employed agent for surgical, endoscopic and laparoscopic interventions [3], [4]. Nevertheless, ICG does not exhibit high specificity for cancer and therefore has not been a preferred approach for surgical oncology. Consequently, attention has been shifted to developing and performing clinical studies using fluorescence agents with sensitivity and specificity to moieties upregulated by cancer, such as receptors or enzymes [5]–[8]. Different targeted fluorescence agents have been granted approval for experimental use in humans, typically under investigational new drug (IND) exceptions [2].

These advances have promoted several developments in fluorescence imaging systems. Different implementations and operational characteristics have been considered, ranging from the wavelengths of operation or spectral channels offered to widely varying detection sensitivity and processing parameters [3]. Linked to them is an emerging need for system benchmarking and quality control, which could further enable establishment of guidelines on good imaging practices [3], [4].

Benchmarking of fluorescence imaging systems can be accomplished using phantoms developed for measuring different operational parameters [9]–[17]. Polyurethane-based phantoms using various constituents for absorption, scattering, and fluorescence have been developed for determining the sensitivity of a fluorescence camera [10], [18], [19]. The performance of fluorescence imaging systems in International System (SI) units of radiance has also been reported based on measurements of a 96-well phantom using titanium dioxide (TiO_2) to simulate tissue scattering and different amounts of quantum dots dispersed in a polyurethane hardener [9]–[20].

We have recently proposed a composite solid phantom for the characterization of multiple system parameters [10]. The phantom enables quantification of system sensitivity as a function of optical properties and depth, optical resolution, diffused fluorescence resolution, magnification, excitation light leakage, parasitic illumination, and color image uniformity [10]. We have used this phantom to compare the performance of two fluorescence imaging systems under various acquisition parameters [21], demonstrating a first attempt to offer a comprehensive characterization of fluorescence imaging systems. However, this initial composite phantom came with limitations: lack of dynamic range assessment and correction of fluorescence image uniformity.

We introduce here a novel composite phantom design that extends the assessment of fluorescence imaging systems, compared to the previously reported composite phantom [10], to include quantification of dynamic range and better determination of the spatial illumination pattern offered by the imaging system. Its novelty, compared to previous phantoms reported by our group [10]–[21] and others [14]–[20], lies not only in the number of performance metrics that it can quantify through one or two acquired frames, but also in its ability to correct the acquired data so that systems of markedly different specifications produce the same readouts from the same field of view, which has not been reported to our knowledge. This is a major step toward high-fidelity fluorescence molecular imaging that would enable comparative multi-center clinical trials, independently of the fluorescence molecular imaging systems employed [22]. We use the novel composite phantom to demonstrate how it can be employed to correct readouts of fluorescence molecular imaging systems and we define a descriptive benchmarking score (BM) that can be employed for rapid system comparisons and/or for quality control. We further show, for the first time, how such a phantom can be employed to assess the performance of a fiber-based fluorescence imaging system, as well as a surgical fluorescence microscope. Finally, we discuss the implication of composite phantoms in quality control and standardization of fluorescence imaging, a process necessary for establishing good imaging practices in clinical environments and for enabling high-fidelity imaging [22].

II. MATERIALS AND METHODS

A. Standardization Phantom

The new phantom proposed (Fig. 1a, b) was built with four main compounds: TiO_2 nanoparticles (Titanium IV Oxide; Sigma Aldrich, St. Louis, MO, USA) for introducing scattering, alcohol-soluble nigrosin (Sigma Aldrich) and bovine hemin ($\geq$90% pure; Sigma Aldrich) for absorption, and organic quantum dots (Qdot 800 ITK, Thermofisher Scientific, Waltham, MA, USA) for fluorescence [10]. Optical properties were quantified as described in [10]. The different features were implemented as indicated in Fig. 1b and are as follows:

Phantom matrix: This provides high scattering and low absorption to simulate a realistic scenario for near-infrared imaging (cyan color, $\mu_s{}' = 22.5$ cm^{-1}, $\mu_a = 0.03$ cm^{-1}).

Sensitivity under different optical properties: This is assessed using nine wells containing QDots of the same concentration but different scattering and absorption (red color). Scattering varies across columns (A: $\mu_s{}' = 10$ cm^{-1}; B: $\mu_s{}' = 7$ cm^{-1}; C: $\mu_s{}' = 3$ cm^{-1}) and absorption across rows (A: $\mu_a = 0.12$ cm^{-1}; B: $\mu_a = 0.25$ cm^{-1}; C: $\mu_a = 0.5$ cm^{-1}).

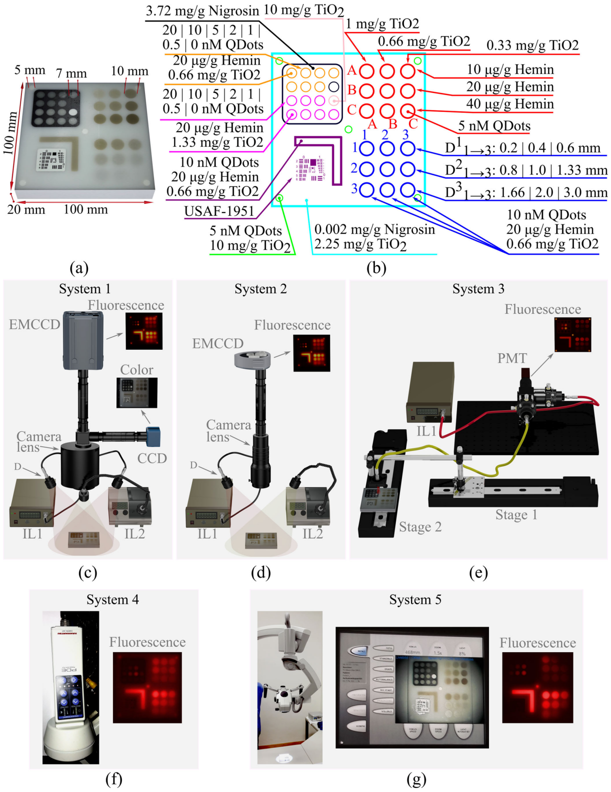


Fig. 1. Standardization phantom and imaging systems. (a) The composite phantom and various dimensions of its structures. (b) The constituents used to build the phantom and their concentration per phantom element. Arrowheads indicate that a group of elements (per row, column, or color) have the constituents indicated by the line tail, while the dotheads indicate the composition of color-coded elements of the phantom. (c) The hybrid fluorescence/color fluorescence imaging system (System 1). (d) The fluorescence imaging system (System 2). (e) The scanning point-based fluorescence imaging system (System 3). (f) The hand-held camera unit of the PDE fluorescence imager (System 4). (g) The camera unit and the articulated arm of the Pentero Surgical Microscope (System 5; left) with the phantom within the field of view (center). A typical fluorescence image of the phantom is depicted on the right. EMCCD: electron multiplying charge-coupled device; CCD: charge-coupled device; IL1: 750 nm excitation source; IL2: field illumination source; D: diffuser; PMT: photomultiplier tube. In (c)–(e), CAD components are courtesy of Thorlabs (www.thorlabs.com).

Sensitivity versus depth: This is assessed using nine wells of identical optical properties at various distances from the phantom's top surface (blue color, $\mu_s{}' = 7$ cm^{-1}, $\mu_s{}' = 0.25$ cm^{-1}).

Resolution: A 1951 United States Air Force chart (USAF-1951) allows assessment of optical resolution, while diffused resolution is assessed by an L-shaped fluorescent structure (deep purple color, $\mu_s{}' = 7$ cm^{-1}, $\mu_a = 0.25$ cm^{-1}).

Dynamic range and light leakage: This novel feature contains a grid of 14 wells with identical optical properties but gradually increasing QDot concentration. The wells are divided into two groups, low-scattering (orange color, $\mu_s{}' = 7$ cm^{-1}) and high-scattering (purple color, $\mu_s{}' = 13$ cm^{-1}), with the absorption identical in the two cases ($\mu_a = 0.25$ cm^{-1}). Both are surrounded by a highly absorbing block (black color block, $\mu_a = 55$ cm^{-1}) to limit diffusion and cross-talk between neighboring wells. This

TABLE I
ACQUISITION SETTINGS OF THE DIFFERENT SYSTEMS

System		Working Distance (mm)	Field of View (cm^2)	Fluence Rate* (mW/cm^2)	Exposure Time (s)	Gain
System 1	f-	300	11.6×11.6	0.9	0.1	3500
	fc-	480	15.0×15.0 13.9×10.2‡	0.7	0.1	3800
System 2		300	10.3×10.3	1.3	0.1	200
System 3		1.5-5.0	point	33 mW**	N/A	10^6
System 4		~300†	21.9×16.4	1.8	N/A	N/A
System 5		467	13.4×10.1	N/A	N/A	13

hand-held; ‡ color camera; * at the phantom surface; ** power at fiber tip.

arrangement can assess the dynamic range of a system with a ratio of minimum to maximum QDot concentration equal to 40. This range is comparable to those reported in other studies of systems with variable bit-depth and/or sensitivity [9]. The fact that our phantom can measure dynamic range for two scattering coefficients is extremely useful for developing and validating quantification procedures [22].

In addition, the top-left quadrant includes one highly absorbing well (black color, $\mu_a = 55\ \text{cm}^{-1}$) and one highly scattering well (pink color, $\mu_s{}' = 100\ \text{cm}^{-1}$) to test the excitation light leakage through the fluorescence optical path, as well as the existence of parasitic illumination due to stray light.

Image uniformity: This is assessed and corrected for using five wells at the phantom's corners and center with relatively high TiO_2 concentration (green color, $\mu_s{}' = 100\ \text{cm}^{-1}$). In the new design, image correction due to excitation light inhomogeneity and/or the presence of the various optical elements is introduced by the QDots present in these wells.

B. Imaging Systems

To demonstrate potential applications of the phantom we employed five fluorescence imaging systems (Table I):

System 1 (Fig. 1c) is a hybrid system developed by our group [21]–[23] that combines near-infrared (NIR) fluorescence, detected with an electron-multiplying charge-coupled device (EMCCD, DV897DCS-BV, Andor Technology, Belfast, UK), and color imaging, performed using a charge-coupled device (CCD; pixelfly qe, PCO AG, Kelheim, Germany). Excitation at 750 nm is achieved using a 300-mW continuous wave (CW) laser diode (BWF2-750-0, B&W Tek, Newark, DE, USA), while field illumination using a 250-W halogen lamp (KL-2500 LCD, Schott AG, Mainz, Germany). System 1 can operate either in fluorescence (*f-System 1*) or in hybrid fluorescence/color mode (*fc-System 1*).

System 2 (Fig. 1d) is a NIR fluorescence molecular imaging system based on an EMCCD (Luca R, Andor Technology) with lower quantum efficiency than System 1 (∼70% for System 1 and ∼40% for System 2 at 800 nm) [21]. Another difference is the band-pass filter employed (D850/40 m, Chroma Technology, Rockingham, VT, USA), which is narrower than the filter in System 1 (ET810/90, Chroma Technology) and operates at wavelengths away from the QDot emission peak (792 nm).

System 3 (Fig. 1e) is a modified version of the intravascular system proposed by our group recently [24]. Two one-dimensional stages (EZSM3D020K, EZ Limo II Series, Oriental Motor Co., Tokyo, Japan) enable the raster-scan of a multi-mode fiber with a core of $200-\mu$m and numerical aperture (NA) of 0.22 (Precision Optics Corporation, Gardner, MA, USA), which serves for both excitation delivery and fluorescence collection. A dichroic mirror (T760lpxr, Chroma Technology) separates excitation from emission light, while the latter is detected by a photomultiplier tube (H5783-20, Hamamatsu Photonics K.K., Shizuoka Pref., Japan). Moreover, long-pass filters (ET780lp, Chroma Technology) and band-pass filters (ET810/90, Chroma Technology) narrow the detection spectrum to a band of 780–855 nm.

System 4 (Fig. 1f) is the CCD-based PDE Near Infrared Fluorescence Imager (Photo Dynamic Eye, Hamamatsu Photonics K.K.) operated in hand-held mode inside the operating room. This is a commercially available system optimized for ICG imaging (ex/em 760/>820 nm) and is clinically employed for fluorescence-guided sentinel node biopsy, esophageal cancer surgery, and other interventions.

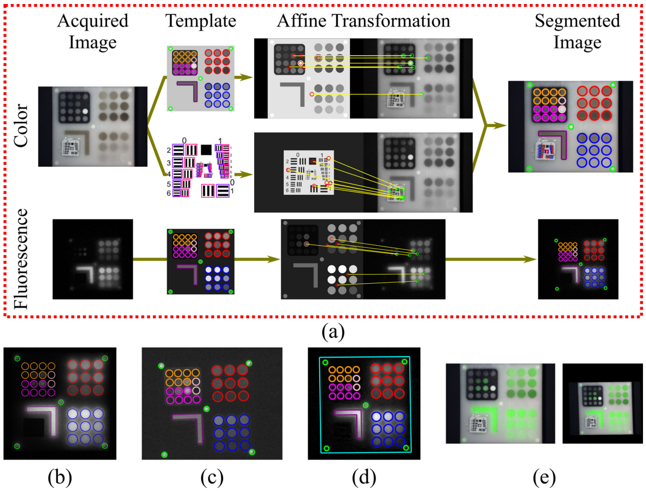
System 5 (Fig. 1g) is the CCD-based Pentero surgical microscope (OPMI Pentero with Infrared and Flow 800, Carl Zeiss Surgical, Oberkochen, Germany). This microscope is routinely used in neurosurgery, plastic and reconstructive procedures, and coronary artery bypass graft surgery. Like System 4, the Pentero surgical microscope is optimized for ICG imaging (ex/em 700–780 / 820–900 nm).

C. Experiments and Data Acquisition

To demonstrate the use of the composite phantom for benchmarking fluorescence imaging systems and/or correcting the acquired data, we performed three experiments.

Experiment 1 employed the composite phantom to assess the pattern of the fluorescence excitation source of 4 different illuminators. We employed f-System 1 with: (a) a single fiber coupled to a circular ground glass diffuser (DG10-220-B, Thorlabs), (b) a bifurcated fiber coupled to one circular ground glass diffuser at each end, (c) a single fiber coupled to a circular tophat engineered diffuser (ED1-C50-MD, Thorlabs), or (d) a bifurcated fiber coupled to a circular tophat engineered diffuser at each end. All illuminators were centered to the field of view and f-System 1 was operated at 1 sec exposure time and 300 mm working distance for all measurements. The camera gain for illuminators a-d was set at, respectively, 500, 1500, 1500, and 2500. The gain was based on the maximum utilization (i.e., between 0% and 80%) of the camera's dynamic range without saturation. Background images (i.e., with excitation source disabled) were also acquired and subtracted from the fluorescence ones.

Experiment 2 employed the composite phantom to correct for the various illumination patters of three markedly different systems, namely Systems 1, 2, and 4. Specifically for System 1 we investigated both operational modes: f-System 1 and fc-System 1. For all measurements, background images were also acquired. Using these data, we further assessed the operational



(a)

(b) (c) (d) (e)

Fig. 2. Phantom image segmentation. (a) The affine transformation between the acquired images and the corresponding templates leads to the segmentation of all phantom components from both color images (top row) and fluorescence images (bottom row). This transformation is extracted by the corresponding points between the templates and the acquired images (seven representative points are shown with the yellow lines for the color image, the USAF-1951 target, and the fluorescence image). Segmentation results shown here are from fc-System 1 and perimeters of all elements have been colored according to Fig. 1b. (b) Phantom segmentation from a fluorescence image acquired by System 2. (c) Phantom segmentation from a fluorescence image reconstructed by the data acquired with System 3. (d) Phantom segmentation from a fluorescence image acquired by System 4. (e) Overlay of fluorescence data onto the color image of the phantom in the fluorescence camera coordinate system (left) or the color camera coordinate system (right).

characteristics of the systems by computing the corresponding BMs (see Section E).

Experiment 3 investigated, for the first time, the use of the phantom to characterize the performance of the raster-scan System 3 and a surgical microscope (System 5). To demonstrate the impact of the working distance for point-based systems on the achieved resolution, System 3 scanned the entire phantom at four different working distances (i.e., 15, 30, 40, and 50 mm). All acquisition settings remained constant between measurements and were as follows: output power at fiber's distal end, 33 mW; scanned region, 600 × 200000 steps; scanning step, 190 × 2$-\mu$m. Performance assessment of System 5 was achieved by imaging the phantom at 467 mm working distance, allowing the entire phantom to fit within the field of view. When the Infrared 800 module was enabled, the field illumination was set at 50% of its maximum value.

D. Data Processing

Following data acquisition, an automated segmentation algorithm was applied to segment the different structures of the phantom (see Fig. 2) and tabulate all measurements in a report file for quick assessment of system performance and benchmarking. This process was based on the speeded-up robust feature (SURF) algorithm, followed by thresholding the distance between features identified in the acquired data and specially designed templates, as previously discussed [21]. The segmentation process is graphically summarized in Fig. 2a, where some of the paired points are depicted. These points define the affine transformation between the acquired data and the corresponding templates. Given this transformation, the segmentation becomes straightforward, as all structures have known coordinates in the templates' coordinate system.

The same segmentation procedure can be applied to any fluorescence system. Fig. 2b shows the segmentation results on a fluorescence image acquired with System 2. The robustness of the method is further demonstrated in Fig. 2c, where the same procedure has been applied to an image reconstructed from raster-scan data acquired by System 3; and in Fig. 2d, where the segmentation has been applied to a fluorescence image captured by System 4.

In the case of hybrid systems with fluorescence and color cameras, the two modalities can be registered using the corresponding geometrical transformations as described in our previous work [21]. Fig 2e depicts this phantom functionality for fc-System 1, either by projecting the fluorescence image onto the color one (left) or vice versa (right). Translation from one imaging plane to another is achieved through the common coordinate system of the templates used, regardless of the individual magnification and/or aspect ratio.

The segmented phantom elements are then used to quantify two statistical metrics: (1) signal-to-noise ratio (SNR), defined as the ratio of the average intensity within a well versus the root mean square noise approximated from the background; and (2) contrast, defined as the ratio of the average intensity within a well minus the average intensity of the background divided by the background. Background is defined as the region of phantom matrix next to the wells.

All data processing was implemented in MATLAB (Mathworks, Natick, MA, USA).

E. Imaging System Benchmarking

A methodology for objective benchmarking was developed by incorporating the metrics afforded by the phantom and is defined by the generic formula:

$$\boldsymbol{BM = sMAPE/N} \tag{1}$$

where $\boldsymbol{sMAPE}$ is the symmetric mean absolute percentage error of each of the 5 metrics quantified (sensitivity *vs* optical properties, sensitivity *vs* depth, resolution, light leakage, and dynamic range). This error is quantified as:

$$\boldsymbol{sMAPE} = \frac{1}{\boldsymbol{n}} \cdot \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \frac{|\boldsymbol{X_i} - \boldsymbol{Y_i}|}{|\boldsymbol{X_i}| + |\boldsymbol{Y_i}|} \tag{2}$$

In (2), $\boldsymbol{n}$ is the number of wells for each metric (i.e., $\boldsymbol{n} = 9$ for the sensitivity *vs* optical properties and *vs* depth), $\boldsymbol{X_i}$ is the average intensity of the $\boldsymbol{i^{th}}$ well ($\boldsymbol{i} = 1, \ldots, \boldsymbol{n}$), and $\boldsymbol{Y_i}$ is the average intensity of the same well that a reference system would record. To define the reference system, we considered readouts to be linear in the wells that assess dynamic range, exponential in the wells that assess sensitivity *vs* depth, and constant across the wells that assess sensitivity *vs* optical properties. The $\boldsymbol{sMAPE}$ of SNR was quantified with respect to a reference of 6 dB and the

$sMAPE$ of contrast was calculated with respect to a reference of 1. Both these reference values represent 95% confidence and are obtained when the readout is two-fold the noise level [21]–[23].

The diffused fluorescence resolution is defined as the shortest line with end points falling on the two edges of the L-shaped structure's concave angle [21]. Based on this definition, it becomes apparent that the ideal system should present diffused resolution equal to zero, as the detected signal is not affected by scattering or absorption. For this reason, to extract the BMs for a real system, the reverse resolution is considered, which is the worst resolution minus the approximated one. The worst resolution corresponds to the hypotenuse of the right angle defined by the two concave edges of the L-shaped element. Thus, division by zero is avoided and increasing ratios correspond to increasing scores. A similar approach was adopted for the optical resolution, quantified by the USAF-1951 target.

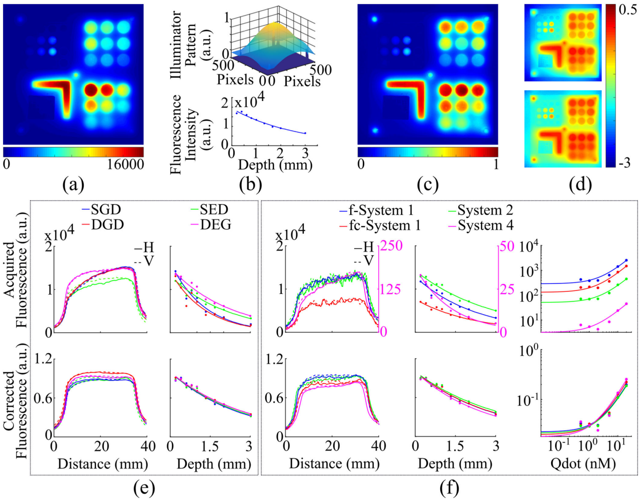


Fig. 3. Fluorescence molecular imaging data correction with the composite phantom. (a) The acquired image from f-System 1 with excitation delivered from illuminator a (see Methods) suffers from inhomogeneity even in structures with uniform distribution of QDot. (b) Interpolating the normalized fluorescence intensity measured at the five wells targeting the illumination correction provided an approximation of the excitation light (top). Following flattening of the excitation source, the average intensity of the nine wells with the QDot distribution at different depths was fitted to an exponential decay (bottom). Its amplitude was employed to normalize the acquired image to the excitation source intensity. (c) The corrected fluorescence image of the phantom. (d) Acquired (top) and corrected (bottom) phantom images depicted on a logarithmic scale. (e) The fluorescence profiles of the fluorescence resolution (left) and depth (right) structures of the phantom obtained using f-System 1 under the four illuminators (see Methods). The corrected fluorescence (bottom) shows better consistency than the acquired one (top). (f) Effects of correction for fluorescence resolution (left), depth (center) and dynamic range under high scattering (right) across different systems. In (e): SGD: single ground diffuser; DGD: double ground diffuser; SED: single engineered diffuser; DEG: double engineered diffuser. In (e)–(f): H: horizontal; V: vertical.

III. Results and Discussion

A. Experiment 1: Assessment and Correction of Fluorescence Excitation Pattern

During Experiment 1 we assessed the spatial patterns of different illuminators (see Methods) using the proposed composite phantom. Fig. 3a shows the fluorescence image acquired under excitation from illuminator a. The image non-uniformity is quite evident, especially in the fluorescence resolution structure at the bottom left quadrant. By employing the five wells located at the corners and center of the phantom (green color in Fig. 1b), the uniformity spatial pattern was approximated (top Fig. 3b). Normalizing with this pattern corrects the acquired image for the intensity distribution, but not for its intensity. The latter was achieved by fitting the 9 wells with different depth (blue color in Fig. 1b) to an exponential function, as seen in Fig. 3b (bottom), after which the acquired data were normalized to its amplitude. Fig. 3c shows the resulting corrected image. Fig. 3d further demonstrates this correction on a logarithmic scale by comparing the acquired (top) and corrected (bottom) images.

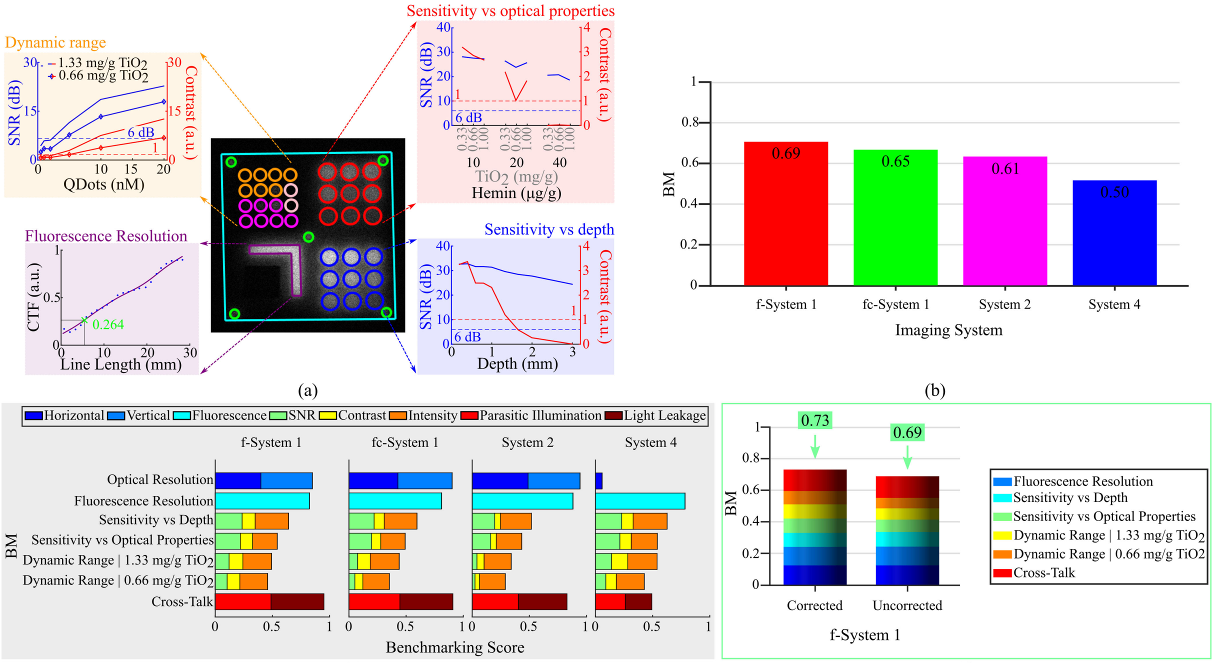
Using the same approach, we corrected all images from the four illuminators. Fig. 3e depicts the comparison results in reference to the fluorescence resolution (left) and depth (right) structures. We found that the corrected data (bottom) present consistent trends among the four different illuminators, while the raw data (top) do not. A representative example is the case of a single ground diffuser (see Fig. 3e). The profile of the fluorescence intensity over the resolution structure is not uniform. However, application of the correction approach compensates for the impact of the fluorescence profile.

Besides excitation light, the observed non-uniformity in the emitted fluorescence could be due to the employed lenses or the excitation-detection geometry. For example, in Fig. 3e we show that even the engineered diffusers did not result in constant intensities, which was caused mainly by the incident angle of the excitation light. Such results demonstrate for the first time that the phantom can be used to perform i) data correction from fluorescence molecular imaging systems and ii) quality control of such systems either during the developmental process or intra-operatively before each imaging session.

B. Experiment 2: Correction and Benchmarking of Imaging Systems

During Experiment 2, results equivalent to those from Experiment 1 were achieved when using the phantom to correct the acquired data from different systems. Fig. 3f shows the fluorescence profile from the structures that test fluorescence resolution (left), depth (center), and dynamic range with high scattering (right) after data acquisition with f-System 1, fc-System 1, System 2, and System 4 (upper row). Following the correction procedures, all wells demonstrate consistent fluorescence emission for all systems (lower row). To the best of our knowledge, this is the first time that fluorescence molecular imaging systems of markedly different specifications demonstrate equivalent response following correction with a phantom. This indicates that referencing of acquired data during clinical trials or longitudinal studies is feasible.

Besides correcting the fluorescence data, the same five wells of the phantom can also be used to correct for non-uniform field illumination in case of hybrid fluorescence/color imaging



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

Fig. 4. Benchmarking of fluorescence imaging systems with markedly different specifications. (a) The fluorescence image of the phantom (center) as acquired by fc-System 1 was segmented and analyzed per quadrant (see Methods). Metrics related to dynamic range, sensitivity (as a function of optical properties and of depth), and fluorescence resolution were quantified per element. (b) A unique BM based on all metrics was calculated for each imaging system used in Experiment 3. (c) Each unique BM was the combination of individual scores per metric. Each score can be decomposed into specific assessment metrics (i.e., SNR or contrast), allowing the in-depth analysis of a system's performance. (d) Benchmarking between corrected and uncorrected f-System 1 with illuminator b and exposure time of 1 s. The BM is the aggregation of each distinct BM.

systems. Such a process does not affect the color information of the acquired data, but instead provides high-quality color images, as we previously demonstrated [21].

Experiment 2, further, aimed to validate the phantom for inter-system benchmarking, by comparing imaging systems differing in acquisition settings (f-System 1 *vs* fc-System 1), or in technical specifications (System 1 *vs* System 2 *vs* System 4). Fig. 4 depicts representative results from this comparison.

Fig. 4a shows the quantified metrics to assess the performance of fc-System 1 after the acquired image had been corrected for image uniformity. These metrics were translated into BMs (see Methods) relative to an "ideal" imaging system. This score, since it is ratiometric, can then be used to compare different imaging systems. For example, Fig. 4b depicts the BMs per system employed in Experiment 2, as derived by (1), following the quantification of the distinct BMs (Fig. 4c). A representative interpretation of those figures is that System 4 showed lower optical resolution and greater cross-talk than all other systems. Indeed in Fig. 4c it is shown that System 4 presents relatively low values in these performance assessment metrics, which further limit its overall performance.

Next, we examined the effect of the correction process on the BMs. Fig. 4d shows the BMs before and after correction for f-System 1 under illuminator b (see Methods, Experiment 1), where all distinct BMs have been stacked to visualize the cumulative BM per system. When compared to the uncorrected system, we see that the correction improved the overall BM (0.73 *vs* 0.69). The non-uniform image, as expected, affected the individual scores and led to an overestimation of the BM. For example, the cross-talk scores are 0.91 *vs* 0.98 for the corrected system and the uncorrected system, respectively, mostly because of the Gaussian profile of the illumination. Contrast, SNR, and dynamic range are similarly affected by non-uniform illumination and/or detection. Algorithmic correction by flat-fielding does not, however, improve these parameters, since signal and noise are both affected. Other approaches should be investigated.

C. Experiment 3: Application to a Raster-Scan System and Surgical Microscope

Experiment 3 focused on application of the phantom to raster-scan (System 3) and narrow-field (System 5) systems. To the best of our knowledge, this is the first time that a phantom designed for wide-field fluorescence molecular imaging is applied to point-based or narrow-field modalities.

Fig. 5a shows the BMs for System 3 at four working distances from 1.5 mm to 5 mm. All scores were nearly the same, yet the system did not perform equally well at all distances. For example, fluorescence resolution score varied from 0.94 for a working distance of 5 mm to 0.99 at 1.5 mm. Fig. 5b shows that this can translate to a resolution difference of almost 1 mm. On the other hand, SNR score was 0.41 at a working distance of 5 mm and 0.25 at 1.5 mm (blue color in Fig. 1b). This is because the shorter the working distance is, the greater is the fraction of excitation light reflected from the sample surface and collected by the fiber, contaminating the acquired signal with

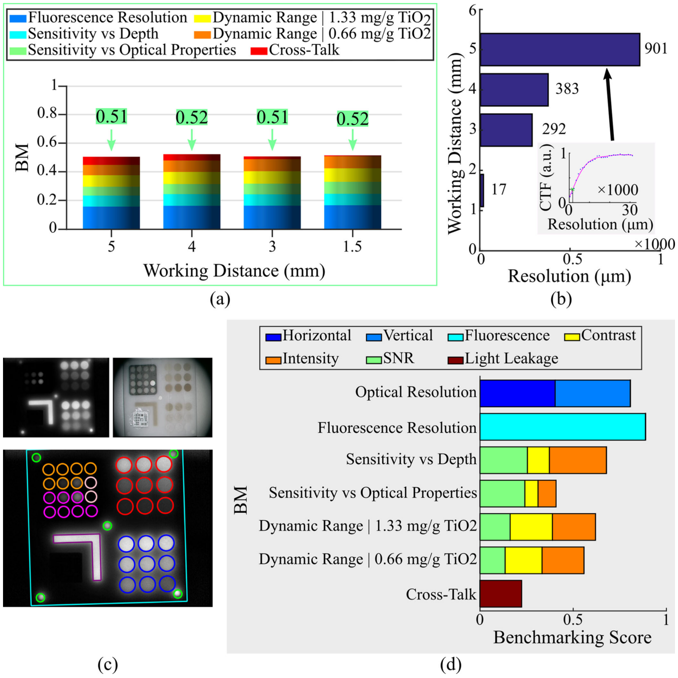


Fig. 5. Validation of phantom-based benchmarking of surgical microscopy and raster scan systems. (a) Individual BMs for raster-scan System 3. (b) Fluorescence resolution as a function of working distance for System 3. The inset shows the contrast transfer function vs resolution at a working distance of 5 mm. (c) Fluorescence (top left) and color (top right) images acquired from System 5. Phantom elements are segmented (bottom) for further analysis. (d) Individual BMs for System 5.

fiber fluorescence. These analyses highlight the complementary information about system performance that can be gained from careful inspection of the aggregate BM as well as the individual metrics.

The data acquired and segmented in Fig. 5c enabled the quantification of the BM of System 5 (Fig. 5d). The BM of this system was quantified at 0.60, which is equivalent to System 2 (Fig. 4b). Nevertheless, the true score of System 5 would be much smaller if correction for the fluorescence induced by the field illumination were applied. However, due to the wide absorption spectrum of the QDots, a significant portion of the emitted fluorescence is induced by the white-light source, making such correction impractical.

IV. DISCUSSION

In the work presented herein we extend the capabilities of a multiparametric phantom that we previously developed and demonstrate that it can support comprehensive correction and benchmarking of a range of prototype and commercially available fluorescence imaging systems, including the hand-held system PDE imager, a surgical microscope and a raster-scan system. This allows, for the first time, the detailed comparison of systems with markedly different specifications and their correction with reference to a composite phantom.

One of the most significant factors determining the operation of a fluorescence imaging system is the spatial independence of fluorescence readout sensitivity over the entire field of view. Several inhomogeneities (i.e., due to excitation illumination or lenses/filters properties) can lead to false negative results, which can translate to erroneous tumor delineation and/or detection during fluorescence imaging-guided surgery. We demonstrate that application of the proposed phantom can correct for such factors that degrade the interpretation of the acquired data (Fig. 3). Despite the different illuminators and/or systems, the applied correction led to consistent results. This has the potential to impact the translational research of fluorescence molecular imaging, as markedly different systems can be referenced to the phantom and thus quantitative comparisons depend only on working distance and tissue optical properties. In addition, the phantom has the potential to serve as a validation target for algorithmic approaches that quantify and/or correct tissue optical properties (i.e., fluorescence cross-talk between wells of different optical properties or depth).

Thus, a unique finding of this study is the ability of the phantom to correct the acquired data for system performance variations, so that markedly different systems provide the same readouts for the same field of view. This is an important prerequisite for achieving high-fidelity fluorescence molecular imaging, as upon system correction the only factors affecting data acquisition are the tissue optical properties and geometry, which can then be corrected for using reversion [22]. Even without reversion, using the phantom for quality control and data correction can allow monitoring of the system's performance stability over the course of a clinical trial; referencing of the acquired data to the same performance characteristics, even if the data were acquired by different systems; and minimizing false positives or negatives due to a system's non-uniform excitation/acquisition profile.

Quantification of dynamic range is another important feature of the phantom. It allows determination of not only the detection limits, but also the resolution for discriminating different fluorophore concentrations. Such metrics are needed in order to interpret fluorescence imaging data reliably, as well as allow data comparisons among the variety of imaging systems available. The phantom presented herein provides this feature and allows assessment of dynamic range under conditions of low and high background scattering. The proposed phantom accomplishes the work of multiple phantoms described in other approaches for quantification of dynamic range [9], [13], [20].

The performance comparison of fluorescence imaging systems is an additional requirement towards the clinical translation of fluorescence molecular imaging and the evaluation of novel fluorescent agents, since it can register the operational characteristics of different systems, which subsequently enables referencing of data acquired by different systems and/or at different sites to one standard. Recently Zhu *et al.* [20] proposed a detailed approach that can accurately quantify various acquisition parameters of an imaging system in absolute SI units. This work can be very useful towards the development of an imaging system, but the complexity in its application limits its use inside the operating room. Furthermore, the exclusion of the excitation subsystems from the analysis limits the applicability to assess different clinically relevant imaging systems for their overall performance. In contrast, the methodology proposed here has the potential to provide equivalent measures in a ratiometric (unit-less) mode and additionally can be employed for correction of readouts and quality control prior to surgical procedures. The overall process, as described here, requires acquisition of at most two images. In addition, it is a non-contact procedure, which eliminates the requirement for phantom sterilization. As proof of concept we demonstrated here the application with System 4 inside the operating room.

Although this work focuses on the correction and benchmarking of wide-field fluorescence imaging systems, we additionally investigated the possibility of using the phantom with other fluorescence modalities, specifically a surgical microscope and a raster-scanning fluorescence system. We show that the proposed technique also has the potential to be employed for these types of systems and to provide useful information regarding their functionality. Ultimately, we aim to develop an equivalent, but much smaller phantom, for use with endoscopes and higher-resolution surgical microscopes. Such systems are frequently used in modern operating rooms.

Future work should explore other potential methods for correcting image uniformity and for calculating individual and overall BMs. In addition, the photostability of the phantom's constituents is well known and documented [9]–[11]. Nevertheless, further work is needed to validate its stability not only over time, but also after extensive exposure to light and temperature or other environmental changes. Additional fluorophores should also be investigated given that the wide excitation spectrum of QDots can contaminate the data with unintended fluorescence due to stray illumination.

V. CONCLUSION

Here we propose a way to standardize fluorescence molecular imaging systems using a phantom that is compatible with ICG, as well as numerous novel near-infrared fluorescent agents still in clinical trials. Extensive testing of the phantom under different illumination schemes and markedly different systems demonstrated its potential for integration into the standard procedures for fluorescence molecular imaging, as well as for use during the development of algorithms for data correction or referencing. The proposed standardization framework may accelerate and expand clinical translation of fluorescence molecular imaging for surgical guidance.

ACKNOWLEDGMENT

The authors would like to thank Constantin Wolff and Lukas Ritschl for the assistance with the Pentero surgical microscope and Dr. A. Chapin Rodríguez for his insightful comments during the preparation of this manuscript.

REFERENCES

[1] W. B. Nagengast *et al.*, "Near-infrared fluorescence molecular endoscopy detects dysplastic oesophageal lesions using topical and systemic tracer of vascular endothelial growth factor A," *Gut*, vol. 68, no. 1, pp. 7–10, 2019.

[2] M. Koch and V. Ntziachristos, "Advancing surgical vision with fluorescence imaging," *Annu. Rev. Med.*, vol. 67, no. 1, pp. 153–164, 2016.

[3] A. V. Dsouza *et al.*, "Review of fluorescence guided surgery systems: identification of key performance capabilities beyond indocyanine green imaging," *J. Biomed. Opt.*, vol. 21, no. 8, 2016, Art. no. 080901.

[4] B. Zhu and E. M. Sevick-Muraca, "A review of performance of near-infrared fluorescence imaging devices used in clinical studies," *Brit. J. Radiol.*, vol. 88, no. 1045, 2015, Art. no. 20140547.

[5] G. M. van Dam *et al.*, "Intraoperative tumor-specific fluorescence imaging in ovarian cancer by folate receptor-[alpha] targeting: First in-human results," *Nat. Med.*, vol. 17, no. 10, pp. 1315–1319, 2011.

[6] L. E. Lamberts *et al.*, "Tumor-specific uptake of fluorescent bevacizumab-IRDye800CW microdosing in patients with primary breast cancer: A phase I feasibility study," *Clin. Cancer Res.*, vol. 23, no. 11, pp. 2730–2741, 2017.

[7] Q. R. J. G. Tummers *et al.*, "Intraoperative imaging of folate receptor alpha positive ovarian and breast cancer using the tumor specific agent EC17," *Oncotarget*, vol. 7, no. 22, pp. 32144–32155, 2016.

[8] N. J. Harlaar *et al.*, "Molecular fluorescence-guided surgery of peritoneal carcinomatosis of colorectal origin: A single-centre feasibility study," *Lancet Gastroenterol. Hepatol.*, vol. 1, no. 4, pp. 283–290, 2016.

[9] B. Zhu, J. C. Rasmussen, and E. M. Sevick-Muraca, "A matter of collection and detection for intraoperative and noninvasive near-infrared fluorescence molecular imaging: To see or not to see?," *Med. Phys.*, vol. 41, no. 2, 2014, Art. no. 022105.

[10] M. Anastasopoulou *et al.*, "Comprehensive phantom for interventional fluorescence molecular imaging," *J. Biomed. Opt.*, vol. 21, no. 9, 2016, Art. no. 091309.

[11] P. Krauter *et al.*, "Optical phantoms with adjustable subdiffusive scattering parameters," *J. Biomed. Opt.*, vol. 20, no. 10, 2015, Art. no. 105008.

[12] M. Roy *et al.*, "Homogenized tissue phantoms for quantitative evaluation of subsurface fluorescence contrast," *J. Biomed. Opt.*, vol. 16, no. 1, 2011, Art. no. 016013.

[13] B. Leh *et al.*, "Optical phantoms with variable properties and geometries for diffuse and fluorescence optical spectroscopy," *J. Biomed. Opt.*, vol. 17, no. 10, 2012, Art. no. 108001.

[14] B. Zhu *et al.*, "Validating the sensitivity and performance of near-infrared fluorescence imaging and tomography devices using a novel solid phantom and measurement approach," *Technol. Cancer Res. Treatment*, vol. 11, no. 1, pp. 95–104, 2012.

[15] Ł. Szyc *et al.*, "Development of a handheld fluorescence imaging camera for intraoperative sentinel lymph node mapping," *J. Biomed. Opt.*, vol. 20, no. 5, 2015, Art. no. 051025.

[16] R. G. Pleijhuis *et al.*, "Near-infrared fluorescence (NIRF) imaging in breast-conserving surgery: Assessing intraoperative techniques in tissue-simulating breast phantoms," *Eur. J. Surg. Oncol.*, vol. 37, no. 1, pp. 32–39, 2011.

[17] A. M. De Grand *et al.*, "Tissue-like phantoms for near-infrared fluorescence imaging system assessment and the training of surgeons," *J. Biomed. Opt.*, vol. 11, no. 1, 2006, Art. no. 014007.

[18] B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," *J. Biomed. Opt.*, vol. 11, no. 4, 2006, Art. no. 041102.

[19] V. Ntziachristos and D. Hyde, "In-vivo fluorescence imaging: applications, future trends & approaches to standardization," in *Standardization and Quality Assurance in Fluorescence Measurements II: Bioanalytical and Biomedical Applications*, U. Resch-Genger, Ed., Berlin, Germany: Springer, 2008, pp. 549–560.

[20] B. Zhu *et al.*, "Determining the performance of fluorescence molecular imaging devices using traceable working standards with SI units of radiance," *IEEE Trans Med Imag.*, vol. 35, no. 3, pp. 802–811, Mar. 2016.

[21] D. Gorpas *et al.*., "Benchmarking of fluorescence cameras through the use of a composite phantom," *J. Biomed. Opt.*, vol. 22, no. 1, 2017, Art. no. 016009.

[22] M. Koch, P. Symvoulidis, and V. Ntziachristos, "Tackling standardization in fluorescence molecular imaging," *Nat. Photon.*, vol. 12, no. 9, pp. 505–515, 2018.

[23] J. Glatz *et al.*., "Concurrent video-rate color and near-infrared fluorescence laparoscopy," *J. Biomed. Opt.*, vol. 18, no. 10, 2013, Art. no. 101302.

[24] D. Bozhko *et al.*., "Quantitative intravascular biological fluorescence-ultrasound imaging of coronary and peripheral arteries in vivo," *Eur. Heart J. Cardiovasc. Imag.*, vol. 18, no. 11, pp. 1253–1261, 2017.