Benchmarking of fluorescence cameras through the use of a composite phantom

Dimitris Gorpas,[a,b] Maximilian Koch,[a,b] Maria Anastasopoulou,[a,b] Uwe Klemm,[b] Vasilis Ntziachristos[a,b,*]

[a] Technical University Munich, Chair for Biological Imaging, Arcisstraße 21, Munich, Germany, D-80333
[b] Helmholtz Zentrum München, Institute for Biological and Medical Imaging, Ingolstädter Landstraße 1, Neuherberg, Germany, D-85764

Abstract. Fluorescence molecular imaging (FMI) has shown potential to detect and delineate cancer during surgery or diagnostic endoscopy. Recent progress on imaging systems has allowed sensitive detection of fluorescent agents even in video rate mode. However, lack of standardization in fluorescence imaging challenges the clinical application of FMI, since the use of different systems may lead to different results from a given study, even when using the same fluorescent agent. In this work we investigate the use of a composite fluorescence phantom, employed as an FMI standard, to offer a comprehensive method for validation and standardization of the performance of different imaging systems. To exclude user interaction, all phantom features are automatically extracted from the acquired epi-illumination color and fluorescence images, using appropriately constructed templates. These features are then employed to characterize the performance and compare different cameras to each other. The proposed method could serve as a framework towards the calibration and benchmarking of fluorescence molecular imaging systems, to facilitate their clinical translation.

Keywords: fluorescence imaging, standardization, rigid phantom, benchmarking.

***Vasilis Ntziachristos, E-mail: v.ntziachristos@tum.de

1 Introduction

An important goal of surgical oncology is the complete removal of malignant tissue. However, limitations in human vision and tactile feedback lead to insufficient intraoperative inspection of tumor margins. Typically, post-surgical assessment of excised specimen with histopathology is required for verification if a complete tumor resection was achieved. Histological analysis of biopsies and surgical specimen is nevertheless time consuming, labor intensive, and subject to sampling errors as involves specialized personnel and subjective interpretation of readouts. Importantly, post-surgical findings of positive tumor margins may require secondary procedures. Thus, new technologies for real-time guidance are needed to address limitations that may lead in incomplete surgical treatment of disease. Indeed, statistics with breast conserving surgeries, for

example, show that the incomplete resection rates range between 20% and 50%[1-3], entailing recurrent surgical intervention with associated risks[4].

Fluorescence imaging has the potential to improve surgical and endoscopic guidance and positively impact the clinical management and prognosis of numerous diseases. Even though indocyanine green (ICG) and other fluorescent agents have been considered for surgical guidance[5-7], including robot assisted interventions[8, 9], their ability for tumor delineation is limited due to lack of specificity[10]. Conversely, fluorescent agents that target specific biochemical and molecular features can improve demarcation of malignant tissue and have received approvals for experimental clinical use[10-12].

A critical issue associated with the clinical translation of fluorescence molecular imaging relates to the reproducibility of the collected measurements. In particular, images acquired from the same target using different fluorescence cameras may vary considerably when the employed systems have markedly different specifications. Methods that standardize fluorescence imaging are therefore becoming necessary for assessing the performance of fluorescence systems and agents and for providing a reference to the recorded data. Consequently, a number of phantoms have been suggested for comparing fluorescence imaging systems[13-16]. Polyurethane-based phantoms were considered recently for assessing the sensitivity of fluorescence cameras or for quantifying the excitation light leakage into the acquired fluorescence images[13, 14]. Tissue mimicking phantoms simulating optical properties and/or geometry of sample tissues have also been applied for training surgeons or for validating new hardware configurations and software methodologies[15, 17, 18]. Nevertheless, most of these phantoms resolve one or a few parameters and do not allow for comprehensive characterization of all variables associated with fluorescence imaging performance.

In this work we sought to identify a methodology that could offer seamless benchmarking of fluorescence cameras. We have recently proposed a composite FMI phantom, i.e. a phantom capable of integrating multiple targets within the field of view of a fluorescence camera, so that multiple camera characteristics can be examined with a single or a few image acquisitions[19]. Using this composite phantom, we develop an approach to process the fluorescence phantom image and quantify different camera performance parameters automatically. We show, for the first time, how the use of a composite phantom can be employed for the comprehensive calibration of a camera system. Further, we demonstrate how composite phantoms can be employed for comparing systems of different specifications. The described benchmarking method may become critical for standardization of imaging systems with broader applications for clinical translation of fluorescence molecular imaging.

2 Materials and Methods

2.1 Standardization phantom

We have previously proposed a rigid phantom for interrogating different aspects of fluorescence and optical imaging performance[19]. The phantom contains a number of imaging targets and resolves different fluorescence features (Fig.1). In particular, as shown in Fig. 1b, each quadrant of the phantom tests different performance parameters, i.e. (i) sensitivity as a function of the optical properties (red color); (ii) sensitivity as a function of the depth from the top surface (blue color); (iii) resolution of the fluorescence and optical imaging (purple color); (iv) cross-talk from the excitation light (pink color). The five wells at the corners and center of the phantom (green color) have been added to assess the field illumination (i.e. illumination for enabling reflectance color imaging) homogeneity when optical measurements are performed through a color camera.

The main constituent of the phantom is a transparent polyurethane (WC-783 A/B, BJB Enterprises, Tustin, USA). The fluorescence compounds are organic quantum dots (Qdot® 800 ITK™, Q21771MP, Thermofisher Scientific Waltham, USA). Scattering is imposed by anatase TiO_2 nanoparticles (Titanium IV Oxide, Sigma Aldrich, St. Louis, USA); absorption is mimicked by alcohol solvable nigrosin (Sigma Aldrich, St. Louis, USA) in the phantom main body and by Hemin (Sigma Aldrich, St. Louis, USA, from bovine ≥ 90%) in the different wells. Selection of the specific phantom materials was based on the stability of their optical properties over time and their ability to create homogeneous mixtures[14, 19]. The specific mixtures content of all phantom elements are given in Fig. 1b. The procedure of the phantom's preparation has been explicitly described elsewhere[19].

2.2 Imaging Systems

To develop a methodology that employs composite phantoms for comparing different systems we used two imaging systems (Fig. 2). The system in Fig. 2a (Camera I) is a modified version of the one that has been developed, characterized, and reported by our group elsewhere[20]. Briefly, a 750 nm CW laser diode (BWF2-750-0, B&W Tek, Newark, Delaware, USA) with a maximum power of 300 mW is used to excite the fluorescence compounds of the phantom, while field illumination is enabled by a 250 W halogen lamp (KL-2500 LCD, Schott AG, Mainz, Germany). The laser power incident to the phantom at a working distance of 15 cm was measured at 85 mW/cm^2, which is lower than the maximum permissive exposure according to American National Standards Institute (ANSI) standard for eye exposure. A short-pass filter (E700SP, Chroma Technology, Rockingham, Vermont, USA) is used to remove the near infrared (NIR) component of the field illumination and thus to eliminate the cross-talk between fluorescence detection and field illumination optical paths (F_1 in Fig. 2a). Ground glass diffusers (DG10-220, Thorlabs, Newton,

New Jersey, USA) are used to achieve uniform illumination of the field of view from both light sources (D in Fig. 2a). The optical signal is collected by a motorized zoom/focus lens (CVO GAZ11569M, Goyo Optical Inc, Asaka, Saitma, Japan) and spectrally resolved in two channels by a dichroic mirror (700DCXXR, AHF analysentechnik AG, Tubingen, Germany) (DM in Fig. 2a). The first channel, which is within the spectral range 720 – 850 nm, is relayed through a NIR achromatic doublet pair (MAP10100100-B, Thorlabs, Newton, New Jersey, USA) (RL_1 in Fig. 2a), filtered by a NIR emission filter (ET810/90, Chroma Technology, Rockingham, Vermont, USA) (F_2 in Fig. 2a) and recorded by an iXon electron multiplying charge-coupled device (DV897DCS-BV EMCCD, Andor Technology, Belfast, Northern Ireland). The second channel, which is within the spectral range 450 – 700 nm, is relayed through a visible achromatic doublet pair (MAP10100100-A, Thorlabs, Newton, New Jersey, USA) (RL_2 in Fig. 2a), filtered by a short-pass filter (ET700SP-2P, Chroma Technology, Rockingham, Vermont, USA) (F_3 in Fig. 2a) and recorded by a 12-bit color CCD camera (pixelfly qe, PCO AG, Kelheim, Germany).

Camera I can operate under two configurations: i) with both cameras (EMCCD and CCD) enabled for simultaneous acquisition of fluorescence and color images (Camera I-FC) or ii) using only the EMCCD camera; thus capturing only fluorescence images (Camera I-F).

The second system (Camera II; Fig.2b) is also based on EMCCD detection (Luca R, Andor Technology, Belfast, Northern Ireland). Camera II has four major differences compared to Camera I: (i) it lacks the color imaging channel (450 – 700 nm spectral band), (ii) it has different operational characteristics (see Table 1), (iii) it uses a different fluorescence filter (D850/40m, Chroma Technology, Rockingham, Vermont, USA) (F_4 in Fig. 2b), and (iv) it employs a different lens (Zoom 7000 Macro Lens, Navitar, New York, USA). The differences between the two fluorescence imaging systems are summarized in Table 1.

Table 1 The main differences between Camera I and Camera II.

	Camera I	**Camera II**
Resolution (pixels)	512×512	1004×1002
Pixel Size (μm)	16×16	8×8
Quantum Efficiency (at 800 nm)	~70%	~40%
Dynamic Range (bit)	16	14
Detection Wavelength (nm)	810/90	850/40
Color Reflectance Imaging	Yes (450 – 700 nm)	N/A

Data acquisition and control of cameras was enabled via the Solis software (Solis I, Andor Technology, Belfast, Northern Ireland) and a GPU based C++ software developed by our group[20]. All data processing was implemented in MATLAB (Mathworks Inc., Massachusetts, USA).

2.3 Data acquisition protocols

We explored the use of the composite phantom in association with two acquisition parameters: 1) the level of pixel-binning of the camera sensor and 2) the working distance, i.e. the distance from the camera lens to the surface of the target object. For all experiments performed, fluorescence was enabled by the same excitation source, that is the 750 nm laser diode (Fig. 2), and images were acquired with room lights turned off. The integration time was set at 0.1 s, to resemble real-time measurements as they are performed in vivo. To ensure minimization of boundary effects, the phantom was placed on top of a highly absorbing material.

Table 2 The acquisition settings for all the investigated cases.

Experimental Configuration	Working Distance (mm)	Binning	Gain	Temperature (°C)
Camera I-F Camera I-FC	320	1×	4000	-70
Camera II-a	320	1×	200	-20
Camera II-2×a	320	2×	200	-20
Camera II-4×a	320	4×	100	-20
Camera II-b	200	1×	200	-20
Camera II-2×b	200	2×	200	-20
Camera II-4×b	200	4×	60	-20

To examine the effects of pixel-binning, both cameras were positioned at the same 320 mm working distance from the phantom surface. This distance is a representative working distance for a wide range of intraoperative applications. Fluorescence images were then acquired with different gains, cooling temperatures and pixel binning, as listed in Table 2.

The influence of the working distance was also examined by operating Camera I at 320 mm distance from the target while changing the working distance of Camera II to 200 mm. Fluorescence images were acquired from Camera II without pixel-binning (Camera II-b), with 2× (Camera II-2×b) and with 4× (Camera II-4×b) pixel-binning. The gain of Camera II-4×b was reduced compared to the other configurations of Camera II in order to avoid saturation. Table 2 summarizes all acquisition settings for both tests implemented.

For every fluorescence image acquired, a corresponding dark image, i.e. image with the excitation light disabled and maintaining constant all other acquisition settings, was also captured. This image was used for estimating the effects of ambient illumination, that is stray light emitted from sources different than the ones of the two systems (i.e. computer monitors, optical mice, indication light emitting diodes on electronic devices, to name a few). Further, subtraction between the fluorescence and dark images compensated for possible parasitic illumination (i.e. ambient illumination at the detection wavelengths) or dark current influence on the validation process.

In addition, reflectance images were acquired for every experimental configuration shown in Table 2. These images were used to estimate the optical parameters of each camera (i.e. magnification and optical resolution, see Sec 2.4 below). The reflectance images, for all configurations of Camera I-C and Camera II, were acquired with room lights turned on. On the other hand, for the configuration of Camera I-FC the reflectance images were acquired by the CCD camera with field illumination enabled and room lights turned off.

2.4 Camera Performance Assessment

To calibrate and compare different imaging systems with minimal user intervention we developed an automated method for the detection of all the composite phantom elements. This method was based on the application of the speeded-up robust features (SURF) algorithm[21] applied on the acquired images and specially designed templates. The distance between the two sets of features was then computed and thresholded based on an efficient approximate nearest neighbor search[22]. The outcome of this process was two sets of image points, one set corresponding to the template and one to the acquired image. These points were then used to estimate the geometric transformation between the two images. Two phantom templates were designed: i) one that was used for the fluorescence images (Fig. 3a), and ii) one that was used for the reflectance images (Fig. 3b).

Employing the geometric transformation derived from the abovementioned process predefined points of interest were projected from the templates onto the acquired images. These points include (i) the four corners of the phantom, (ii) the center and one perimeter point of all the circular phantom elements, (iii) the six corners of the L-shaped phantom element, and (iv) the four corners of the USAF-1951 target, as well as the four corners of all the target's line elements. These points are adequate to extract all phantom components and consequently to quantify camera performance metrics, i.e. magnification, optical resolution, diffused fluorescence resolution, excitation light leakage and parasitic illumination, sensitivity and field illumination homogeneity as described in the following:

2.4.1. Magnification

By detecting the boundaries of the phantom in the acquired images, the magnification, $\boldsymbol{M}$, is approximated by the following formula:

$$\boldsymbol{M} = \frac{\boldsymbol{D}_{\text{phantom}}}{\sqrt{\boldsymbol{D}_{\boldsymbol{pw}}^2 \cdot \left(\boldsymbol{u}_1 - \boldsymbol{u}_2\right)^2 + \boldsymbol{D}_{\boldsymbol{ph}}^2 \cdot \left(\boldsymbol{v}_1 - \boldsymbol{v}_2\right)^2}}, \tag{1}$$

where $\boldsymbol{D}_{\text{phantom}} = 100$ mm is the phantom's width, and the denominator represents the Euclidean distance in Cartesian coordinates between two adjacent corners of the phantom on the imaging plane. In Eq. (1) $\left(\boldsymbol{u}_1, \boldsymbol{v}_1\right)$ and $\left(\boldsymbol{u}_2, \boldsymbol{v}_2\right)$ are the pixel coordinates of the two phantom corners, $\boldsymbol{D}_{\boldsymbol{pw}}$ is the pixel width, and $\boldsymbol{D}_{\boldsymbol{ph}}$ the pixel height. This metric is then expressed as the well-known magnification ratio $1:\boldsymbol{M}$. The magnification can be approximated either from the acquired fluorescence images or from the reflectance ones.

2.4.2. Optical resolution

The optical resolution is approximated through the USAF-1951 target, which consists of a series of elements (having two sets of three lines separated by spaces equal to one line's width) at right angles, organized in groups. This metric is quantified as cycles per mm (c/mm) through the general formula:

$$\boldsymbol{F} = 2^{\text{group} + (\text{element} - 1)/6}, \tag{2}$$

which provides the frequency, $\boldsymbol{F}$, of any target element in any target group. In addition, the line or space width of each element can be estimated from the frequency as $\boldsymbol{L}_{\boldsymbol{O}} = \left(2 \cdot \boldsymbol{F}\right)^{-1}$ in mm.

For each one of the bounding boxes of the USAF-1951 target elements, as defined by the four corners of the corresponding line, the contrast transfer function (CTF) is quantified by the Michelson's formula[23]:

$$\boldsymbol{CTF}_{\boldsymbol{i}} = \frac{\max\left(\boldsymbol{I}_{\boldsymbol{i}}\right) - \min\left(\boldsymbol{I}_{\boldsymbol{i}}\right)}{\max\left(\boldsymbol{I}_{\boldsymbol{i}}\right) + \min\left(\boldsymbol{I}_{\boldsymbol{i}}\right)}, \tag{3}$$

where I_i denotes the intensity values of the pixels inside the i^{th} bounding box. Adopting the Rayleigh criterion for optical imaging, the limit where CTF is ~26.4%[20] defines the USAF-1951 target's elements that can be fully resolved by the system. The frequency of these elements is then estimated by Eq. (2), and thus the resolution of the system for the specific working distance and field of view is quantified.

Optical resolution can be validated only on reflectance images, acquired either with room lights turned on (i.e. for all Camera II configurations and Camera I-F) or under field illumination (i.e. for Camera I-FC configuration), as described in Sec. 2.3.

2.4.3. Diffused fluorescence resolution

The diffused fluorescence resolution is quantified by the L-shaped element of the phantom. The determination of diffused resolution of a fluorescence imaging system is achieved for a specific set of optical properties of the fluorescent target and the surrounding medium and changes if the optical properties of either one change. Thus, the purpose of the L-shaped element in the phantom is the quantification of resolution improvements that may be offered by a system and/or method.

As in the optical resolution, the CTF is employed for the approximation of the diffused fluorescence resolution. Specifically, the CTF is calculated over every line segment that is perpendicular to the bisector of the L-shaped element (Fig. 3c). In order to exclude any possible bias by the USAF-1951 target, all lines crossing it are not considered. In addition, to remove high frequency noise from the CTF measurements, the CTF distribution across the lines is fitted to a second order polynomial. Limiting the CTF to ~26.4%, the length of the last line, L_F, over this threshold is transformed from image to scene (phantom) coordinates, using the pixel dimensions, $D_{pw} \times D_{ph}$, and the magnification, M:

$$L_F = M \cdot \sqrt{D_{pw}^2 \cdot (u_1 - u_2)^2 + D_{ph}^2 \cdot (v_1 - v_2)^2}\,, \tag{4}$$

where (u_1, v_1) and (u_2, v_2) are the two edges of the line segment. This distance is reported as the diffused fluorescence resolution of the investigated camera.

2.4.4. Excitation light leakage and parasitic illumination

The presence of excitation light leakage through the filters is assessed by the highly scattering and the highly absorbing wells in the upper left quadrant of the phantom (pink color in Fig. 1b). Specifically, the transmission ratio, $R_{exc} = S/N$, is estimated by assuming the average pixel counts (i.e. digitized pixel intensity) from the highly scattering element as the signal component, S, and the average pixel counts from the highly absorbing element as noise, N. If this ratio is equal to or smaller than 1, it implies reduced excitation light leakage, whereas higher values indicate the presence of excitation light in the acquired data.

Parasitic illumination is assessed following the exact same process as for light leakage, but on the dark images. With the excitation source disabled, the highly scattering and absorbing elements of the phantom can only be visualized under the presence of parasitic illumination. The transmission ratio, R_{par}, will provide a quantified indicator on the influence and significance of parasitic illumination. This quantity may be related to filter parameters and environmental conditions, such as ambient light, and can be further employed to optimize the system or the lighting conditions.

2.4.5. Sensitivity vs. optical properties and depth

The sensitivity of a fluorescence camera for various optical properties (i.e. different absorption and scattering) or depth has been validated based on readouts from the 18 corresponding phantom wells (red and blue colors in Fig. 1b).

The signal-to-noise ratio (SNR) for each of these wells is approximated through the formula:

$$\boldsymbol{SNR_{dB}} = 20 \cdot \log\left(\frac{\boldsymbol{S}}{\boldsymbol{RMSN}} \right) \tag{5}$$

where $\boldsymbol{S}$ is the average intensity within each well of the raw fluorescence image and $\boldsymbol{RMSN}$ corresponds to the root mean square noise approximated from the phantoms' main body (i.e. excluding all elements). In this way the SNR is normalized for variations that are introduced by the base constituents of the phantom and excitation light leakage. Moreover, considering normal signal distributions, a measurement is assumed to present 95% confidence if the signal is twofold the noise level[20], which in the dB scale corresponds to 6 dB.

The capability of the system to image the fluorescence wells is further assessed through the contrast as expressed by the Weber's fraction, $\boldsymbol{C}$ [24]:

$$\boldsymbol{C} = \frac{\boldsymbol{S} - \boldsymbol{N}}{\boldsymbol{N}} \tag{6}$$

where $\boldsymbol{N}$ is the average pixel counts from the highly scattering element of the phantom (right pink color in Fig. 1b). The higher this metric for a fluorescence element of the phantom, the better a fluorescence camera can discriminate it from light leakage. Similar to SNR, 95% confidence is achieved when the signal is twofold the noise level, which corresponds to $\boldsymbol{C} = 1$. As expected from the definitions of SNR and contrast, these quantities are figures of merit for benchmarking different systems as a function of the specific phantom, rather than absolute values to characterize a single system.

2.4.6. Field illumination homogeneity

The homogeneity of the field illumination can be approximated using the 5 highly scattering elements located at the 4 corners and the center of the phantom (green color in Fig. 1b). A normalized average intensity equal to 1 for all 5 elements indicates a homogenous illumination profile, whereas the higher the deviation between the 5 values, the higher is the illumination heterogeneity. We note that while this phantom offers a crude spatial sampling of the illumination field, integration of more reflecting elements could be considered for more accurate determination of field illumination.

Flat-fielding can be achieved by applying bicubic splines interpolation of the average intensity from each of the 5 elements and dividing the acquired reflectance images with the resulting profile[19].

2.5 Image Registration

In addition to the performance assessment of fluorescence cameras, the phantom enables interrogation of the registration between fluorescence and color images, as it applies to fluorescence/color imaging systems, such as Camera I. The transformation from the image coordinates of color camera, $(\boldsymbol{u}_c, \boldsymbol{v}_c)$, to the coordinates of the fluorescence camera, $(\boldsymbol{u}_f, \boldsymbol{v}_f)$, is performed through the linear system:

$$\begin{bmatrix} \boldsymbol{u}_c \\ \boldsymbol{v}_c \\ 1 \end{bmatrix} = \mathbf{A} \cdot \begin{bmatrix} \boldsymbol{u}_f \\ \boldsymbol{v}_f \\ 1 \end{bmatrix} \tag{7}$$

The 3×3 affine transformation $\mathbf{A}$ can be estimated by Eq. (7) after solving the correspondence problem, i.e. the extraction of points of interest visible by both imaging modalities. Nevertheless, the fiduciary markers of the phantom (see Sec. 2.4) can serve as points of interest and thus can be

used to solve the correspondence problem and through that approximate the affine transformation $\mathbf{A}$.

3 Results

Fig. 4 shows the fluorescence and reflectance image pairs as acquired by three experimental configurations of the two cameras employed in this study. Specifically, Fig.4a shows the fluorescence image acquired by configuration Camera I-F and Fig. 4b depicts the corresponding reflectance image acquired with the room lights turned on. The equivalent image pair acquired by configuration Camera II-a is shown in Fig. 4c and Fig. 4d. Finally, for the experimental configuration Camera I-FC the fluorescence image of the phantom is shown in Fig. 4e and the reflectance image acquired by the color camera with enabled the field illumination and the room lights turned off is shown in Fig. 4f. The different sensor size and pixel resolution between the EMCCD and CCD of Camera I lead to the different aspect ratios between panels (e) and (f). The fluorescence images shown in Fig. 4 are employed for the determination of the fluorescence performance parameters of the two cameras for the various experimental configurations (i.e. fluorescence resolution, excitation light leakage, parasitic illumination, and sensitivity), while the reflectance images are used for the determination of the optical parameters (i.e. optical resolution, magnification).

The boundaries of the phantom elements shown in all panels of Fig. 4 were determined by the projection of the points of interest describing each element from the templates to the phantom images via the application of the SURF algorithm (see Sec. 2.4). These elements are employed for the automated quantification of the camera parameters without any user intervention.

3.1 Assessment of Fluorescence Imaging Sensitivity

Fig. 5 and Fig. 6 demonstrate the fluorescence imaging performance of the first and third configurations in Table 2. Fig. 5 depicts the fluorescence measurements from configuration Camera I-F. Specifically, Fig. 5a shows the pixel counts of the 9 wells with different optical properties and Fig. 5b,c depict the corresponding SNRs and contrast $(\boldsymbol{C})$ achieved as estimated by Eq. (5) and (6) respectively. The visualization scheme adopted for the SNR and contrast is based on the corresponding thresholds (i.e. 6 dB for SNR and 1 for $\boldsymbol{C}$). This provides a rapid visual assessment whether these thresholds are exceeded in any of the wells. Further, the height of the saturated area, that is the red part of the cylinder, provides information regarding the order of magnitude that the achieved metric is higher than the corresponding threshold. The same metrics as derived from the 9 wells with different depths from the phantom's top surface are shown in Fig. 5d though Fig. 5f. Fig.6 shows the corresponding measurements for configuration Camera II-a in Table 2, and in the same order as in Fig.5.

Two systems can be considered of similar sensitivity if 1) they succeed to pass the SNR and $\boldsymbol{C}$ thresholds in the same number of wells, and 2) the quantified SNR and $\boldsymbol{C}$ values are of equivalent distances from the corresponding thresholds. For the two systems considered in this study, it is evident from Fig. 5 and Fig. 6 that Camera I-F outperforms Camera II-a in all sensitivity metrics: (i) the pixel counts of Camera I-F exceed tenfold those from Camera II-a; (ii) Camera I-F passes the 6 dB SNR threshold at 8 out of the 9 wells with different optical properties and at all depths, while Camera II-a passes the SNR threshold at 6 of the wells with different optical properties and at 7 wells of different depth; (iii) although $\boldsymbol{C}$ for both camera configurations is above the threshold for all wells, Camera I-F presents twice as stronger contrast than Camera II-a.

Fig. 7 depicts the summary of quantitative comparisons between Camera I-F and all Camera II experimental configurations (see Table 2). Fig. 7a depicts the pixel counts from the 9 wells with the different optical properties, while Fig. 7b,c show the corresponding achieved SNRs and $\boldsymbol{C}$. Similarly, panels (d) through (f) show the pixel counts, SNRs and $\boldsymbol{C}$ from the 9 wells with the different depths from the phantom's top surface.

Panels (a), (b), (d), and (e) of Fig. 7 show that binning or reduction of the working distance for Camera II improves both pixel counts and SNRs for all wells. For example, in Fig. 7a there is more than tenfold increase in the pixel counts between Camera II-a and Camera II-4×b for the well containing 20 μg/g Hemin, 0.66 mg/g TiO_2, and 10 nM QDots (the right well of the B group in Fig. 1b), which is translated to almost 50% SNR increase as seen in Fig. 7b. In the last two columns of Table 3 the achieved sensitivity of all cameras is given by means of minimum (i.e. the minimum value above the 6 dB threshold) and maximum SNR for the 9 wells with different optical properties, $SNR_{OP,min}$ and $SNR_{OP,max}$ correspondingly, and the ones with different depth from the top surface of the phantom, $SNR_{D,min}$ and $SNR_{D,max}$, correspondingly. However, the observable improvement in sensitivity comes with contrast reduction, as seen in panels (c) and (f) of Fig. 7. This is due to the increase of the excitation light leakage and the parasitic illumination (Table 3). Although Camera II-a presents inferior performance compared to Camera I-F, the measurements shown in Fig. 7 reveal that modification of its acquisition parameters (i.e. working distance and/or binning) can eventually lead to comparable to Camera I-F performance.

3.2 Assessment of optical/fluorescence parameters

Table 3 tabulates the magnification $(\boldsymbol{M})$, the optical $(\boldsymbol{L_R})$ and diffused $(\boldsymbol{L_F})$ resolution, and the excitation light leakage $(\boldsymbol{R_{exc}})$ from both cameras and all experimental configurations as

quantified by the images of Fig. 4. Parasitic illumination $\left(\boldsymbol{R}_{par}\right)$ quantified by the transmission ratio from the corresponding dark images, is also included. The magnification does not deviate much from one measurement to another since the phantom was covering most of the field of view for all configurations.

Table 3 Quantification of various validation metrics from both cameras under all experimental configurations.

Camera	$\boldsymbol{M}$	$\boldsymbol{L_R}$ (mm)***	$\boldsymbol{L_F}$ (mm)	$\boldsymbol{R_{exc}}$	$\boldsymbol{R_{par}}$	$SNR_{OP,min}$ $SNR_{D,min}$ (dB)****	$SNR_{OP,max}$ $SNR_{D,max}$ (dB)*****
Camera I-F	1:13.3	0.13 × 0.10	1.3	1.9	1.3	6.5_{C1} 9.4_{D9}	18.0_{B10} 17.0_{D1}
Camera I-FC	1:15.9* 1:17.5**	0.04 × 0.04**	1.9	1.8	1.9	6.5_{B1} 6.8_{D6}	14.0_{B10} 13.4_{D1}
Camera II-a	1:14.2	0.05 × 0.05	1.4	1.3	1.1	7.4_{A10} 6.2_{D6}	10.9_{B10} 11.4_{D1}
Camera II-2×a	1:14.2	0.11 × 0.10	3.0	1.4	1.3	6.5_{C1} 7.1_{D9}	14.6_{B10} 15.4_{D1}
Camera II-4×a	1:14.2	0.25 × 0.20	3.7	1.5	1.6	9.0_{C1} 8.8_{D9}	16.2_{B10} 17.0_{D1}
Camera II-b	1:13.8	0.06 × 0.04	2.2	1.7	1.1	7.4_{A10} 8.2_{D6}	12.9_{B10} 14.2_{D1}
Camera II-2×b	1:13.6	0.14 × 0.10	3.6	1.8	1.2	6.4_{B1} 7.3_{D8}	15.0_{B10} 16.2_{D1}
Camera II-4×b	1:13.6	0.25 × 0.28	3.7	1.9	1.5	8.0_{C1} 6.0_{D9}	15.0_{B10} 17.0_{D1}

* Fluorescence; ** Color; *** Horizontal × Vertical resolution, **** Indices represent the wells labeling according to Fig. 1b.

The optical resolution depends on the pixel size of each camera and the magnification (i.e. 512×512 pixels for the EMCCD of Camera I and 1002×1004 pixels for Camera II at 16×16 μm and 8×8 μm respectively). Camera I-F, having the largest pixel dimensions, presents lower resolution compared to Camera II-a and Camera II-b, but has similar resolution with Camera II-2×a and Camera II-2×b. The diffused fluorescence resolution depends on the specific optical properties of the fluorescence target and is introduced to interrogate devices that may account for photon diffusion; a function that is not enabled in the experimental study herein.

A limited amount of excitation light leakage in both cameras was observed, as shown in Table 3. At the same working distance, Camera I-F displays higher light leakage than Camera II-a; possibly due to differences of the filters employed. Nevertheless, the light leakage becomes

comparable between the two systems when the working distance of Camera II is reduced to 200 mm. The latter is expected and is due to the reduced distance between the imaged surface, the camera lens and the excitation source. When comparing Camera I-F and Camera II-a it becomes apparent that the level of parasitic illumination is slightly higher in Camera I-F; also possibly due to the different band-pass filters employed in the two cameras. Under binning, however, Camera II presents equal or even higher (i.e. binning 4×) parasitic light contamination than Camera I-F. Finally, the sensitivity measurements shown in the last two columns of Table 3 summarize the findings discussed in the previous section (Sec 2.1) when comparing Camera I with Camera II. Although magnification is substantially different than in all other camera configurations, Camera I-FC shows higher sensitivity to many Camera II acquisition settings.

To identify which of the acquisition settings of Camera II investigated in this study brings the performance of that camera closer to the performance of Camera I, we adopted a least squares method between all metrics quantified through the phantom. This analysis identified Camera II-2×b as the one that approaches better the performance of Camera I.

3.3 Correction of Field Illumination Homogeneity

Fig. 8a depicts the normalized intensity inside the 5 highly scattering wells (green color in Fig. 1b) imaged by the color camera of the Camera I-FC. These values are used to approximate the field illumination profile onto the phantom surface. For the specific measurement, as seen in Fig. 8b, this profile is strongly inhomogeneous. This observation is further confirmed by Fig. 8c, where the phantom image is affected by vignetting. The applied flat-fielding, however, corrected for this inhomogeneity, while preserving all the color information in the acquired image, as it can be seen in Fig. 8d.

3.4 Registration of Color and Fluorescence Images

Fig. 9 shows two versions of the same color image, after flat-fielding, augmented with the corresponding fluorescence image. In panel (a) the fluorescence image has been transformed to the coordinates system of the color image, while panel (b) depicts the inverse transformation (i.e. the color image transformed to the coordinates system of the fluorescence image). These transformations do not require increased computational cost, as all the fiducial markers, required for the estimation of the affine transformation between the two imaging planes, are available through the SURF algorithm. The derived transformation matrices can be employed to provide highly accurate co-registration between the color and fluorescence images.

4 Discussion

The absence of robust methods for benchmarking fluorescence molecular imaging platforms and standardizing fluorescence medical imaging may impose hurdles on its diagnostic and clinical translation. The study herein introduced a framework for the use of composite fluorescence phantoms for the a) automatic characterization of the performance of different imaging cameras and b) calibration of two imaging platforms. The procedure can also employ the calibration parameters extracted to configure different systems so that they offer comparable performance, which may be of relevance to multi-center clinical trials or for producing measurements using different systems. Equilibrating images could be also performed in retrospect by image processing using the calibration parameters extracted by the different systems.

To benchmark two fluorescence imaging systems of markedly different specifications we employed a composite phantom, designed to exhibit a variety of optical and fluorescence characteristics[19]. Fluorescence images from this phantom were acquired using two cameras and different operational modes. The methodology developed provides the means for non-biased

validation of systems by automatically extracting imaging features and thus excluding human interference, using previously developed composite phantoms[19]. Compared to simple phantoms[13-16], composite phantoms can provide a comprehensive characterization of fluorescence imaging system performance using a single or a few images, possibly even at a single snapshot.

As seen in Fig. 4, there is always a relatively large number of phantom elements visible in the acquired images, making the application of the SURF based registration algorithm quite robust and accurate. Nevertheless, other registration algorithms can also be applied with potentially equivalent results, as the novelty of the proposed approach is the use of templates, which even for the case of the fluorescence images, where not all wells are visible, ensures successful extraction of all phantom elements.

The results presented in this study allowed the comparison of the two exemplary cameras employed herein for demonstration purposes and clearly indicated a superiority of Camera I compared to Camera II when operating at equivalent settings (i.e. working distance, pixel-binning, or gain, to name a few). However, modification of one or more of these acquisition parameters impacts imaging performance. Indeed, our results clearly demonstrate that appropriate adjustment of Camera II acquisition parameters can modify the performance and optimize certain performance features to resemble Camera I performance.

Automatic extraction of camera specifications using composite phantoms can by employed in basic research and clinical studies to validate a camera's performance prior to surgery or endoscopy (e.g. under miniaturization of the phantom design) and to confirm selected configurations and proper operation. Automatic feature extraction visualized in composite phantoms can therefore serve on a trouble-shooting role over time and before each experimental process[20]. Another possible use is in quality control during a camera manufacturing or selection

process. Finally, an intended application for the methodology developed is the generation of "equivalency" of data acquired during multi-center clinical trials; where the camera systems employed can be calibrated on the same composite phantoms for referencing purposes in association with the data collected.

Future work includes the development of composite phantoms that can capture a larger number of camera parameters, such as characterizing the dynamic range and spectral response, and to offer a more accurate correction for inhomogeneous illumination. The base material of the phantom described herein presents a relatively high absorbance. This minimizes any possible cross-talk between neighboring wells and photon diffusion is relatively limited. Nevertheless, such implementation is not ideal for validating the performance of systems designed to encounter for photon diffusion. In order to enable such feature, we plan to develop phantoms that will incorporate a number of wells within highly absorbing base material and thus cross-talks between neighboring elements will be eliminated, whereas other wells will be within highly scattering base material, and thus allow for validation of the aforementioned systems. Functions for the automatic extraction of these additional features can then be developed to streamline the detection and analysis of a larger set of calibration parameters. Although there exists evidence in literature regarding the stability of the employed materials[14, 19], we further plan to perform systematic constancy and mechanical integrity tests to validate the stability of the optical properties of the various phantom elements at different environments.

The proposed herein methodology represents an early attempt of standardizing imaging measurements or systems for fluorescence molecular imaging. Overall, we expect that the field of standardization will play a major role in the growth of fluorescence molecular imaging.

Disclosure

All authors declare no conflict of interest.

Acknowledgments

The authors would like to thank Dr. Pilar Beatriz Garcia-Allende for her contribution during the preparation of the phantom and Dr. Ovsepyan Saak for his insightful comments during the preparation of this manuscript. The research leading to these results has received funding by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich-824 (SFB-824), subproject A1 and from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 644373 (PRISAR).

References

1. L. E. McCahill et al., "Variability in reexcision following breast conservation surgery," *J. Am. Med. Assoc.* **307**(5), 467-475 (2012) [doi:10.1001/jama.2012.43].
2. J. F. Waljee et al., "Predictors of Re-excision among Women Undergoing Breast-Conserving Surgery for Cancer," *Ann. Surg. Oncol.* **15**(5), 1297-1303 (2008) [doi:10.1245/s10434-007-9777-x].
3. C. Cellini et al., "Factors Associated With Residual Breast Cancer After Re-excision for Close or Positive Margins," *Ann. Surg. Oncol.* **11**(10), 915-920 (2004) [doi:10.1245/ASO.2004.12.037].
4. C. Chiappa et al., "Surgical margins in breast conservation," *Int. J. Surg.* **11**(Sup1), S69-S72 (2013) [doi:10.1016/S1743-9191(13)60021-7].
5. L. Boni et al., "Clinical applications of indocyanine green (ICG) enhanced fluorescence in laparoscopic surgery," *Surg. Endosc.* **29**(7), 2046-2055 (2015) [doi:10.1007/s00464-014-3895-x].
6. M. Plante et al., "Sentinel node mapping with indocyanine green and endoscopic near-infrared fluorescence imaging in endometrial cancer. A pilot study and review of the literature," *Gynecol. Oncol.* **137**(3), 443-447 (2015) [doi:10.1016/j.ygyno.2015.03.004].
7. T. Yamamichi et al., "Clinical application of indocyanine green (ICG) fluorescent imaging of hepatoblastoma," *J. Pediatr. Surg.* **50**(5), 833-836 (2015) [doi:10.1016/j.jpedsurg.2015.01.014].
8. M. A. Bjurlin, T. R. McClintock, and M. D. Stifelman, "Near-Infrared Fluorescence Imaging with Intraoperative Administration of Indocyanine Green for Robotic Partial Nephrectomy," *Curr. Urol. Rep.* **16**(4), 1-7 (2015) [doi:10.1007/s11934-015-0495-9].

9. D. Daskalaki et al., "Fluorescence in robotic surgery," *J. Surg. Oncol.* **112**(3), 250-256 (2015) [doi:10.1002/jso.23910].
10. M. V. Marshall et al., "Near-Infrared Fluorescence Imaging in Humans with Indocyanine Green: A Review and Update," *Open Surg. Oncol. J.* **2**(2), 12-25 (2010) [doi:10.2174/1876504101002010012].
11. 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.* **17**(10), 1315-1319 (2011) [doi:10.1038/nm.2472].
12. M. J. Whitley et al., "A mouse-human phase 1 co-clinical trial of a protease-activated fluorescent probe for imaging cancer," *Sci. Transl. Med.* **8**(320), 320ra324 (2016) [doi:10.1126/scitranslmed.aad0293].
13. 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.* **41**(2), 022105 (2014) [doi:10.1118/1.4862514].
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. Treat.* **11**(1), 95-104 (2012) [doi:10.7785/tcrt.2012.500238].
15. B. Leh et al., "Optical phantoms with variable properties and geometries for diffuse and fluorescence optical spectroscopy," *J. Biomed. Opt.* **17**(10), 108001 (2012) [doi:10.1117/1.JBO.17.10.108001].
16. B. W. Pogue, and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," *J. Biomed. Opt.* **11**(4), 041102 (2006) [doi:10.1117/1.2335429].
17. Ł. Szyc et al., "Development of a handheld fluorescence imaging camera for intraoperative sentinel lymph node mapping," *J. Biomed. Opt.* **20**(5), 051025 (2015) [doi:10.1117/1.JBO.20.5.051025].
18. 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.* **37**(1), 32-39 (2011) [doi:10.1016/j.ejso.2010.10.006].
19. M. Anastasopoulou et al., "Comprehensive phantom for interventional fluorescence molecular imaging," *J. Biomed. Opt.* **21**(9), 091309 (2016) [doi:10.1117/1.JBO.21.9.091309].
20. J. Glatz et al., "Concurrent video-rate color and near-infrared fluorescence laparoscopy," *J. Biomed. Opt.* **18**(10), 101302 (2013) [doi:10.1117/1.JBO.18.10.101302].
21. H. Bay et al., "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Und.* **110**(3), 346-359 (2008) [doi:10.1016/j.cviu.2007.09.014].
22. M. Muja, and D. G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration," *International Conference on Computer Vision Theory and Applications* 331-340 (2009).
23. A. A. Michelson, *Studies in optics*, Dover Publications, Inc., Mineola, N.Y., United States (1995).
24. E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A* **7**(10), 2032-2040 (1990) [doi:10.1364/JOSAA.7.002032].

Figures List

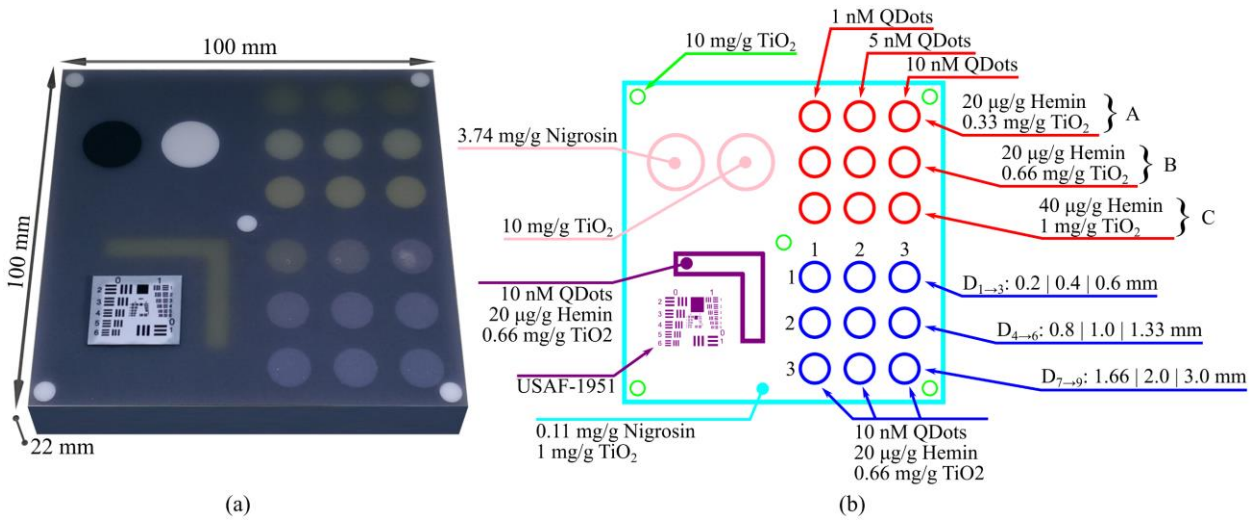


Fig. 1 The standardization phantom employed in this study. (a) The designed phantom and its dimensions. (b) The different compartments per element and/or group of elements of the phantom. The base material is transparent rigid polyurethane. In (b): Arrowheads indicate that a group of elements (per row, column, or color code) have the same constituents, while the dotheads indicate the composition of a single element of the phantom. Color codes: Red - sensitivity vs. optical properties (three sets of optical properties, A, B, and C); Blue - sensitivity vs. depth (index in the 3×3 matrix defines the depth); Purple - resolution; Pink - cross-talk; Green – field illumination homogeneity; Cyan - phantom body; Di - depth from the phantom's top surface.

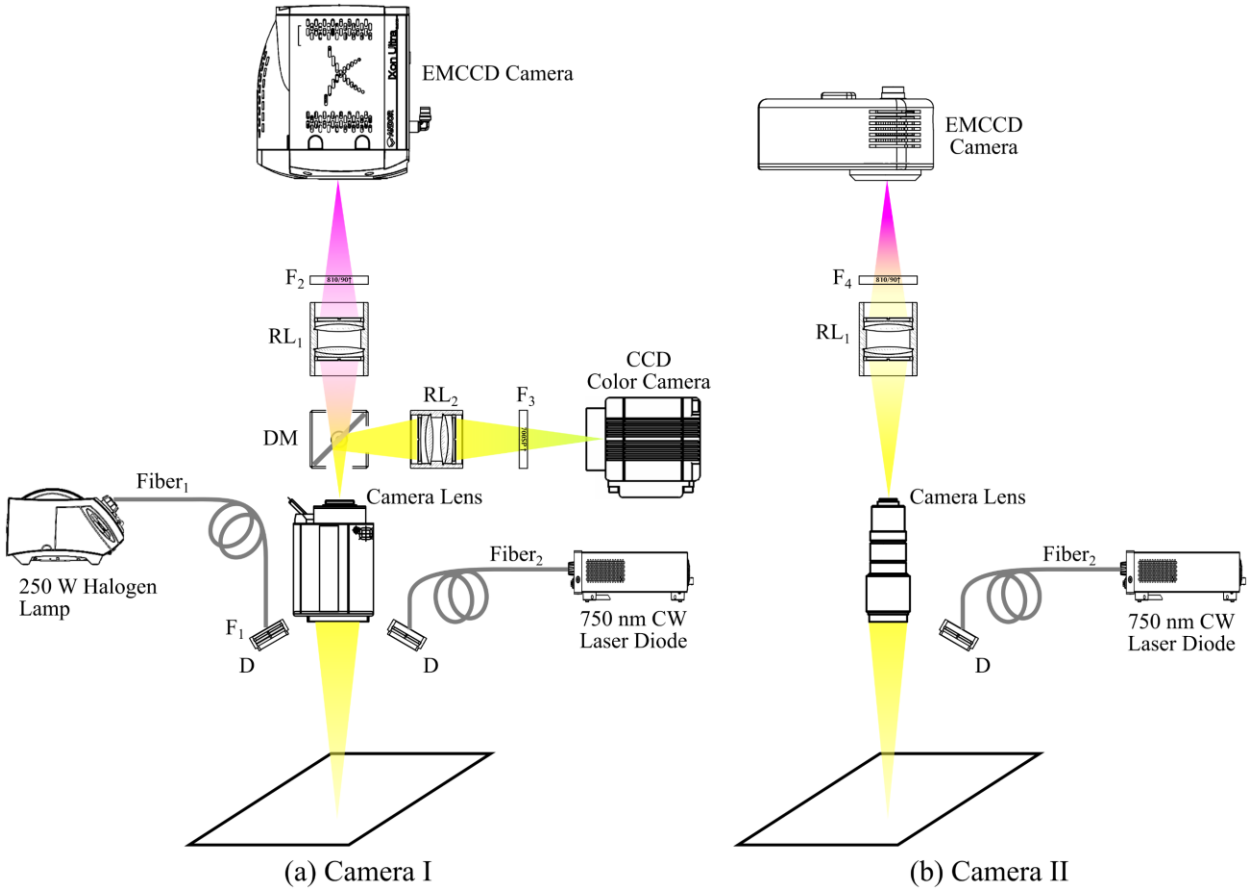


Fig. 2 Schematics of the two imaging systems employed in this study. (a) The fluorescence/color camera (Camera I) and (b) the fluorescence camera (Camera II) were validated and benchmarked through imaging the composite phantom. D: diffuser; F: filter; DM: dichroic mirror; RL: relay lens.

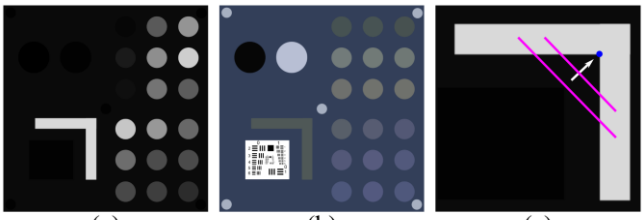


Fig. 3 The templates used for the automated detection of all phantom elements. (a) The template for the fluorescence images includes only the elements of the phantom containing the QDots. (b) The template for the reflectance images, on the other hand, includes all elements of the phantom. (c) The diffused fluorescence resolution is defined as the smallest line perpendicular to the bisector of the L-shaped element (purple line) that can resolve the two edges of the concave vertex (blue dot). The scanning direction of the purple line is indicated by the white arrow and following the template matching can be applied to the fluorescence images.

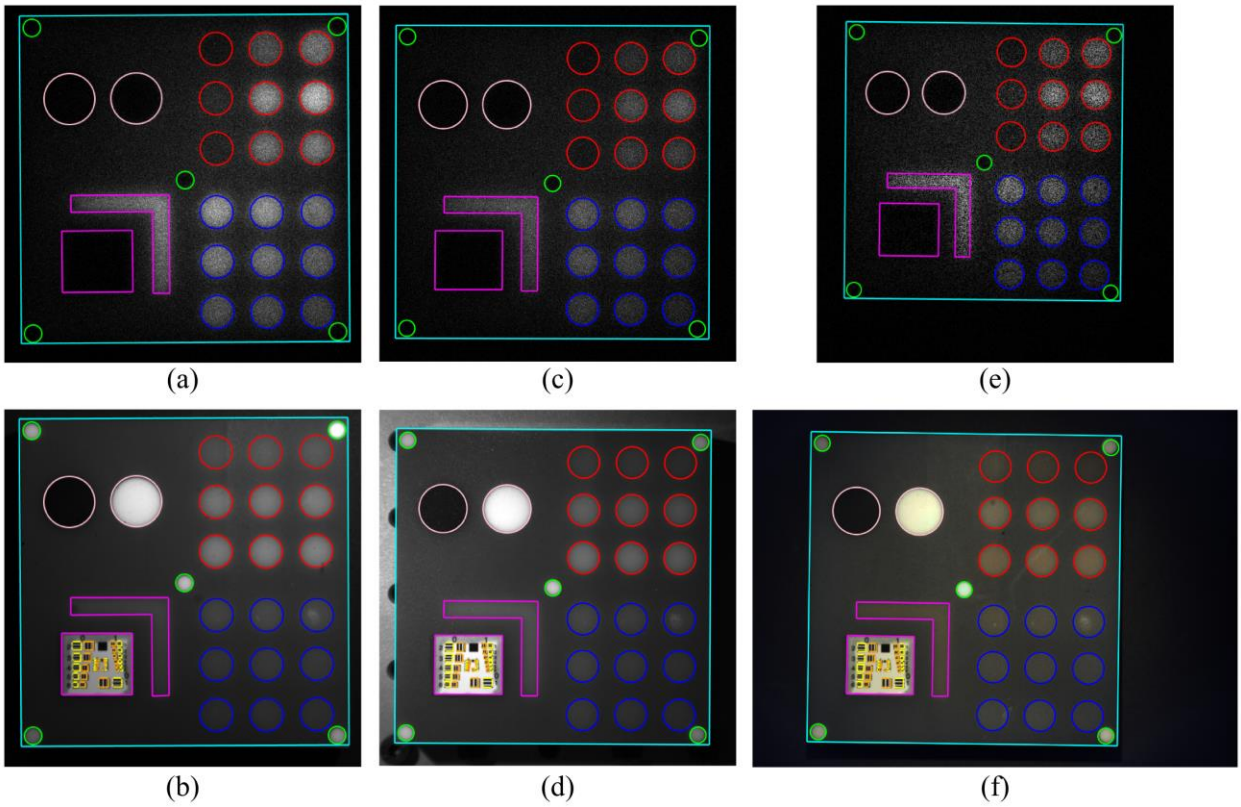


Fig. 4 The fluorescence and reflectance image pairs acquired by three experimental configurations of the two cameras employed in this study, with the phantom's elements highlighted by color-coded edges. (a) The fluorescence image acquired by Camera I-F and (b) the corresponding reflectance image acquired with room lights turned on. (c) The fluorescence image acquired by Camera II-a and (d) the corresponding reflectance image acquired with room lights turned on. (e) The fluorescence image acquired by Camera I-FC and (f) the reflectance image simultaneously acquired by the color camera of Camera I with the field illumination enabled and the room lights turned off. The color-code of the phantom elements correspond to the one of Fig. 1b.

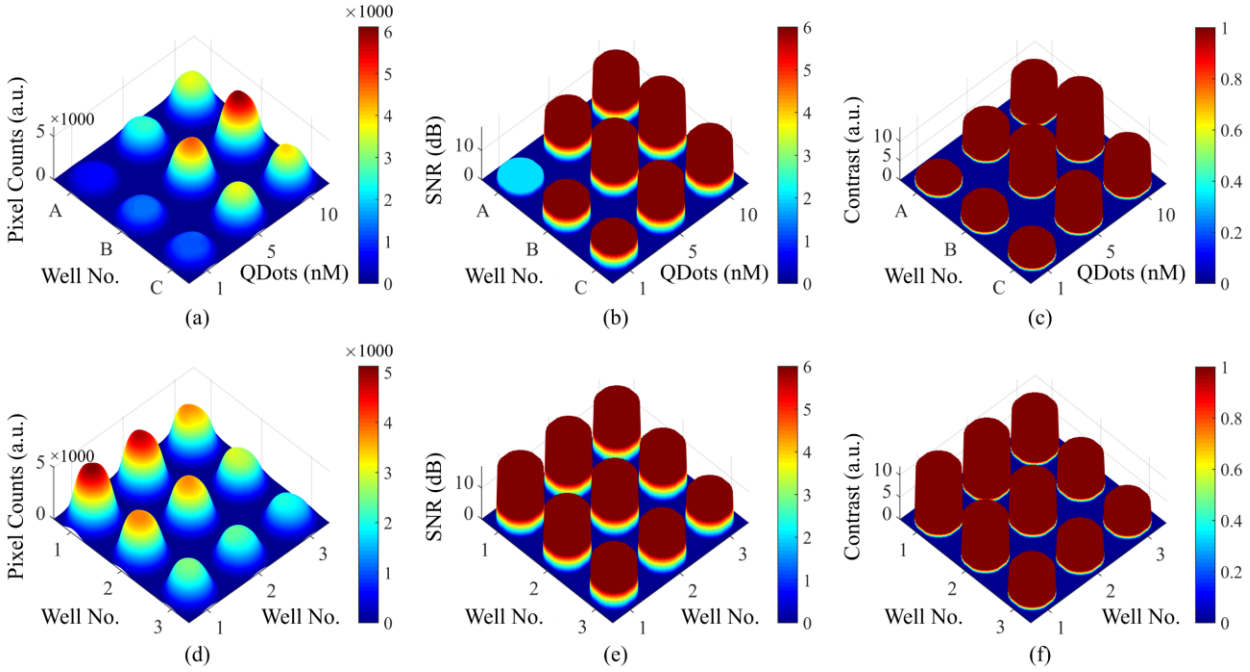


Fig. 5 The fluorescence measurements from experimental configuration of Camera I-F. (a) The pixel counts of the 9 wells with different optical properties. (b) The corresponding SNR and (c) contrast achieved. (d) The pixel counts of the 9 wells with different depth from the phantom's top surface. (e) The corresponding SNR and (c) contrast achieved. SNR and contrast were quantified through Eq. (5) and (6) respectively. Colorbars correspond to the z-axis of each panel and for SNR and contrast metrics define the threshold levels.

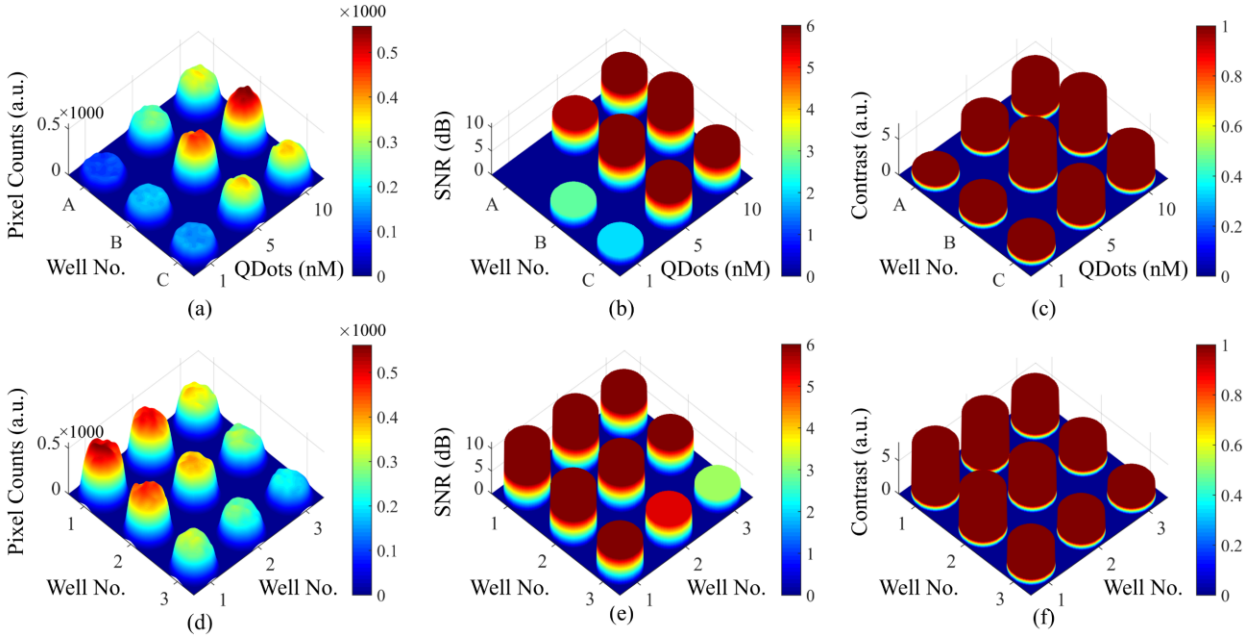


Fig. 6 The fluorescence measurements from experimental configuration of Camera II-a. (a) The pixel counts of the 9 wells with different optical properties. (b) The corresponding SNR and (c) contrast achieved. (d) The pixel counts of the 9 wells with different depth from the phantom's top surface. (e) The corresponding SNR and (c) contrast achieved. SNR and contrast were quantified through Eq. (5) and (6) respectively. Colorbars correspond to the z-axis of each panel and for SNR and contrast metrics define the threshold levels.

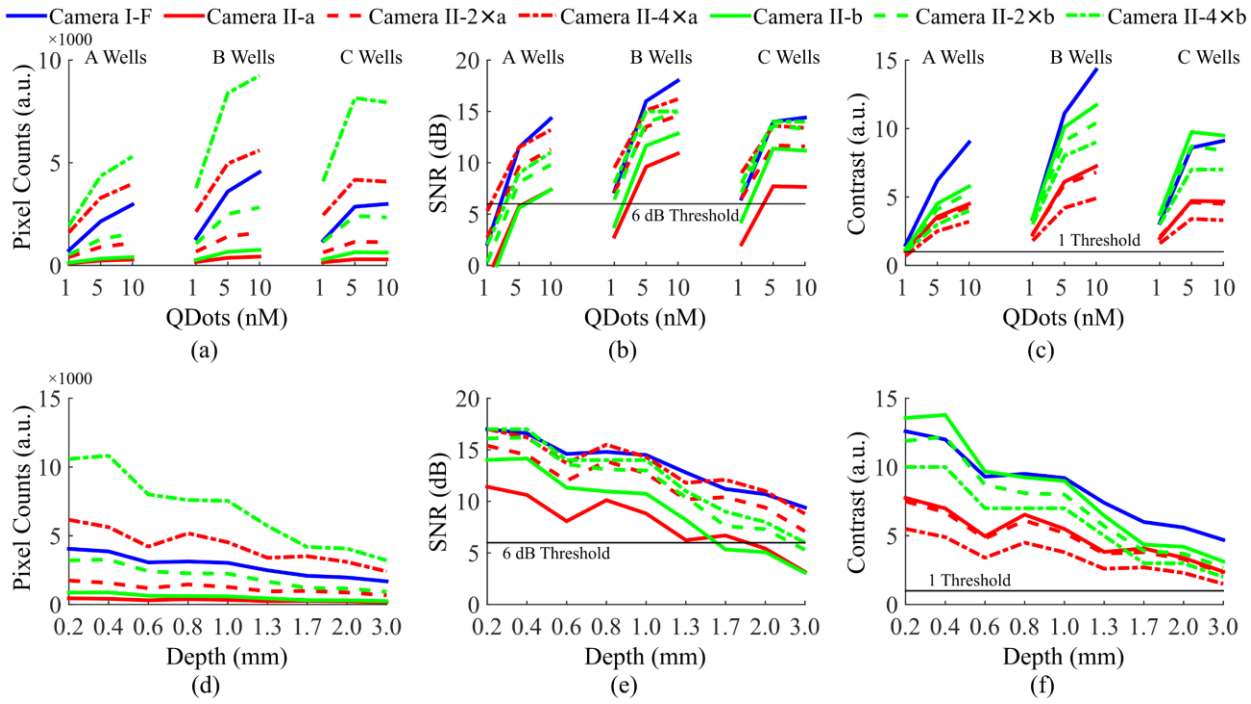


Fig. 7 Quantitative comparisons between Camera I-F and all configurations of Camera II. (a) Pixel counts, (b) SNR, and (c) contrast from the 9 wells with different optical properties. (d) Pixel counts, (e) SNR, and (f) contrast from the 9 wells with different depths from the phantom's top surface. In all panels x-axis labeling corresponds to the labeling of phantom elements shown in Fig. 1b.

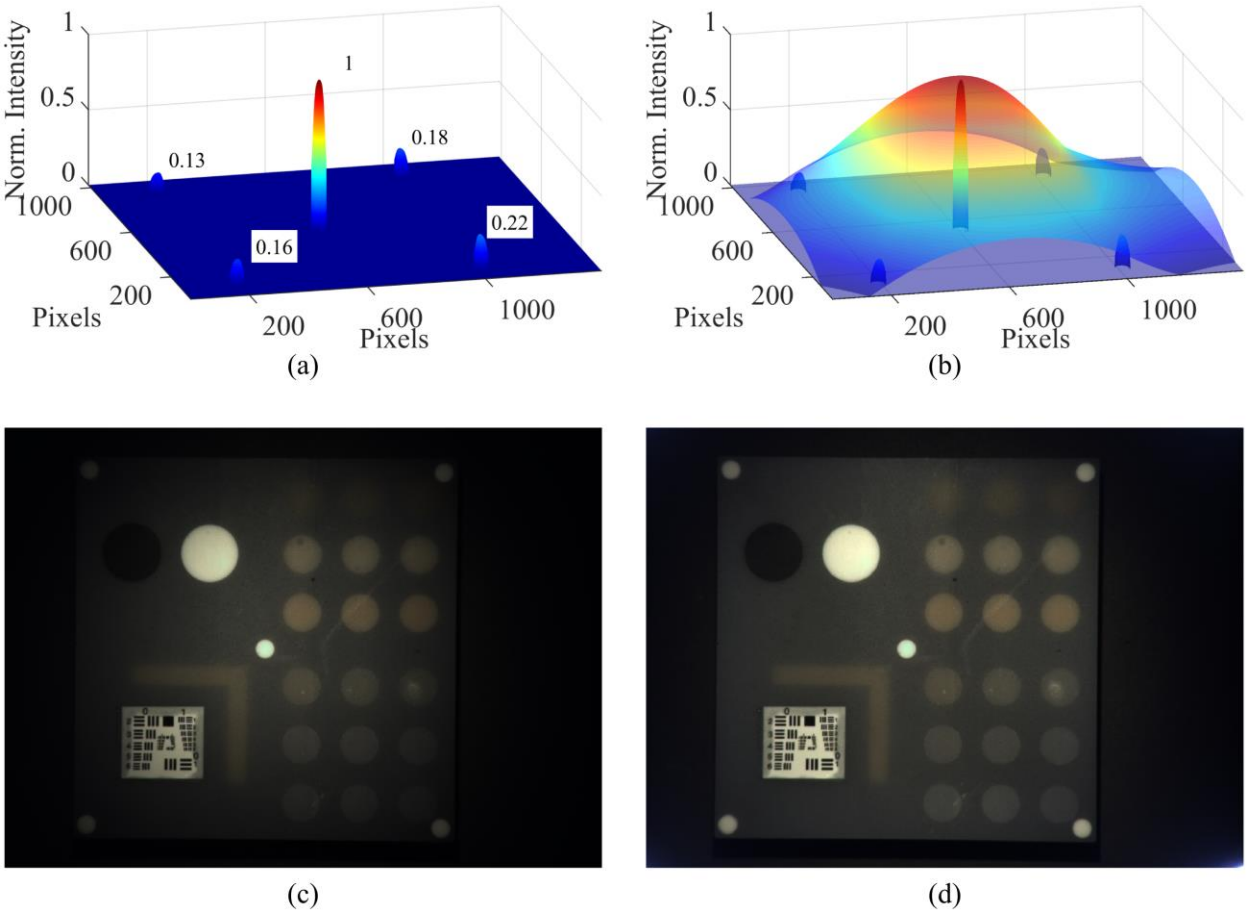
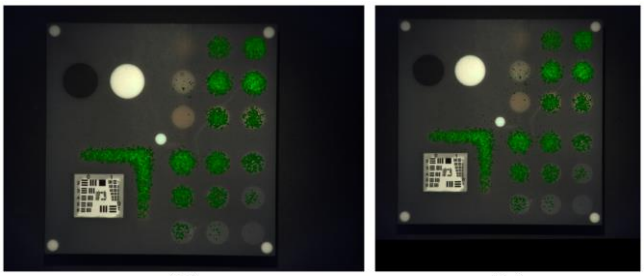


Fig. 8 Field illumination correction through reflectance measurements. (a) The normalized intensity of the 5 highly scattering wells of the phantom. (b) The field illumination profile over the entire field of view. (c) An acquired image from the color camera of Camera I affected by vignetting. (d) Application of flat-fielding to correct for inhomogeneous field illumination.



(a) (b)

Fig. 9 A color image of the phantom augmented with fluorescence data as acquired by Camera I-FC. (a) The augmented image in its original coordinates system. (b) The augmented image transformed to the coordinates system of the fluorescence image.

Table 1 The main differences between Camera I and Camera II.

Table 2 The acquisition settings for all the investigated cases.

Table 3 Quantification of various validation metrics from both cameras under all experimental configurations.