Designing microarray phantoms for hyperspectral imaging validation

Matthew L. Clarke,1,2 Ji Youn Lee,1 Daniel V. Samarov,3 David W. Allen,4 Maritoni Litorja,4 Ralph Nossal,5 and Jeeseong Hwang1,*
1Radiation and Biomolecular Physics Division, Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
2Currently with the National Gallery of Art, Washington, DC, 20565, USA
3Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
4Sensor Science Division, Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
5Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 USA
*jeeseong.hwang@nist.gov

Abstract: The design and fabrication of custom-tailored microarrays for use as phantoms in the characterization of hyperspectral imaging systems is described. Corresponding analysis methods for biologically relevant samples are also discussed. An image-based phantom design was used to program a microarrayer robot to print prescribed mixtures of dyes onto microscope slides. The resulting arrays were imaged by a hyperspectral imaging microscope. The shape of the spots results in significant scattering signals, which can be used to test image analysis algorithms. Separation of the scattering signals allowed elucidation of individual dye spectra. In addition, spectral fitting of the absorbance spectra of complex dye mixtures was performed in order to determine local dye concentrations. Such microarray phantoms provide a robust testing platform for comparisons of hyperspectral imaging acquisition and analysis methods.

© 2012 Optical Society of America

OCIS codes: (110.4234) Multispectral and hyperspectral imaging; (120.0120) Instrumentation, measurement, and metrology; (170.0170) Medical optics and biotechnology; (170.3880) Medical and biological imaging; (170.0180) Microscopy; (350.4800) Optical standards and testing.

References and links
1. M. B. Sinclair, J. A. Timlin, D. M. Haaland, and M. Werner-Washburne, “Design, construction, characterization, and application of a hyperspectral microarray scanner,” Appl. Opt. 43(10), 2079–2088 (2004).
2. C. Glasenapp, W. Mönch, H. Krause, and H. Zappe, “Biochip reader with dynamic holographic excitation and hyperspectral fluorescence detection,” J. Biomed. Opt. 12(1), 014038 (2007).
3. M. L. Huebschman, R. A. Schultz, and H. R. Garner, “Characteristics and capabilities of the hyperspectral imaging microscope,” IEEE Eng. Med. Biol. Mag. 21(4), 104–117 (2002).
4. F. Erfurth, A. Tretjakov, B. Nyuuki, G. Mrotzek, W. D. Schmidt, D. Fassler, and H. P. Saluz, “Two-laser, large-field hyperspectral microarray scanner for the analysis of multicolor microarrays,” Anal. Chem. 80(20), 7706–7713 (2008).
5. J. Y. Lee, M. L. Clarke, F. Tokumasu, J. F. Lesoine, D. W. Allen, R. Chang, M. Litorja, and J. Hwang, “Absorption-based hyperspectral imaging and analysis of single erythrocytes,” IEEE J. Sel. Topics Quantum Electron. 17(6) 1801–1840 (2011).
6. R. Richards-Kortum and E. Sevick-Muraca, “Quantitative optical spectroscopy for tissue diagnosis,” Annu. Rev. Phys. Chem. 47(1), 555–606 (1996).
7. I. Itzkan, L. Qiu, H. Fang, M. M. Zaman, E. Vitkin, I. C. Ghiran, S. Salahuddin, M. Modell, C. Andersson, L. M. Kimerer, P. B. Cipolloni, K. H. Lim, S. D. Freedman, I. Bigio, B. P. Sachs, E. B. Hanlon, and L. T. Perelman, “Confocal light absorption and scattering spectroscopic microscopy monitors organelles in live cells with no exogenous labels,” Proc. Natl. Acad. Sci. U.S.A. 104(44), 17255–17260 (2007).
1. Introduction

The design and characterization of analysis methods for hyperspectral imaging (HSI) is often complicated by the need for appropriate test samples. Even if target signatures are known, the preparation of separate “pure” spectral signals may not be feasible. Indeed, for most scenarios, the detected hyperspectral signals will rarely originate from a single source. For example, a chemical absorption signature will be convolved with light scattering, noise, and instrumental response functions. Therefore, the preparation of phantoms containing known chemical compositions with set spatial profiles would allow for the robust characterization of appropriate experimental parameters and statistical analysis methods.

Printed microarrays offer the potential for high-throughput analysis, and are often employed in nucleotide and protein assays. Generally, a microarray reader records one to four fluorescence channels, with spectral bandwidth determined by optical filters, and compares the background to the signal fluorescence. Replacing the filter-based detectors with a hyperspectral imaging system has been shown to identify and eliminate sources of error, such as spatially localized contaminants [1]. Background fluorescence can be removed by post-processing, or reduced via structured illumination of the sample [2]. Additionally, the narrow bandwidths and numerous channels of hyperspectral imagers allow resolution of fluorophores with similar emission maxima when compared to optical filters [3]. For example, Erfurth et al. [4] calculated a cross-talk of typically <1% between a series of fluorophores after hyperspectral processing. Our ultimate goal is to use hyperspectral imaging techniques for cell and tissue imaging [5]. Here, we demonstrate that microarrays can be expanded beyond fluorescence microscopy of in vitro samples to serve as tools for characterizing hyperspectral absorbance or reflectance measurements for complex, spatially distributed biological systems.

A key challenge in optical spectroscopy and microscopy of tissues lies in decoupling their absorption and scattering signals [6]. Depending on the particular system, the isolation of one such signal may be sufficient, or both signals may need to be considered together to determine
pathological relevance. Multimodal imaging, combining light scattering and absorption microscopy, have been used to study organelles within single cells during apoptosis without the need for exogenous stains [7]. Additional medical studies include the detection of skin cancers in mice [8] and the visualization of tumor hypoxia [9]. Non-invasive imaging facilitates the monitoring of clinical treatments. For example, skin tissue oxygenation levels have been monitored during nitric oxide inhalation by patients [10] and induced vasodilation in the feet of diabetics [11] were followed using reflectance hyperspectral imaging. Coupling hyperspectral instruments to fundus cameras allowed for ophthalmological information to be obtained – notably retinal hemoglobin oxygen saturation levels [12,13].

Indeed, reflectance or absorbance hyperspectral imaging is well suited to the measurement of hemoglobin oxygenation saturation. Yet, several complicating factors may come into play when investigating the analysis of blood oxygenation. Due to the difficulties in preparing appropriate standards, such HSI systems are typically calibrated only at fully saturated and desaturated levels [9]. While hemoglobin and deoxygenated hemoglobin account for the majority of the spectral signal, their spectral signatures extensively overlap. Additionally, other chromophores in the sample may confound the analysis. Any statistical analysis of the data set from such a sample must either determine the spectral signatures and amounts of such interference, or have been sufficiently tested in order to determine the significance of any error introduced by ignoring these factors.

There are several advantages to using a fully controllable system, where various interfering signals can be introduced. However, phantoms designed to mimic biological materials must be carefully considered, as a tradeoff may exist between their manageability and their biological relevance. Typically, it is not feasible to isolate a lone spectral signature, and any model must therefore allow for the presence of overlapping signals. Here, we examine the preparation of microarray phantoms that allows for spatial control of sample placement and a wide range of spectral profiles. These arrays have been analyzed using a hyperspectral imaging system capable of measuring biological samples [5]. We demonstrate the ability to prepare array phantoms containing complex hyperspectral signatures, and detail the initial analysis of these arrays as statistical benchmarks. Specifically, we present the printing of spot arrays consisting of single chromophore dyes, or mixtures of dyes, and show how they may be measured by hyperspectral absorption imaging. Our model system allows for exceptional flexibility in the preparation of a large number of samples. By imaging diverse arrays of known chromophore concentrations (or other appropriate samples), the robustness of the analysis techniques can be determined. Additionally, the number of interfering signals can be easily adjusted. This technique can be readily generalized to the printing of other chromophores or fluorophores, including biological materials such as proteins containing distinct spectral signatures.

2. Experimental methods

Water soluble dyes were selected for the printing based on their absorption profiles, spanning the visible range. Dyes were purchased from Sigma-Aldrich. In particular, the following four dyes were used: acid red 1 (AR), new coccine (NC), erioglaucine (EG), and brilliant blue R (BBR). Two sample types were employed. In the first, three dyes (AR, EG, BBR) having significantly different absorption spectra were mixed in various amounts and printed. In the second, two dyes with similar absorption spectra (AR and NC) were mixed and printed. Dyes were dissolved in water to create stock solutions that provided absorption spectra of similar magnitude (w/w: 4.9% AR; 7.0% NC; 5.1% BBR; 1.9% EG). Stock solutions were then mixed and combined with 75% polyethylene glycol (PEG, 600 amu) to create a dye solution in water containing 50% PEG. The PEG solution ensures high quality print uniformity, allowing for spots of similar size and preventing dye crystallization [14]. These solutions then were loaded into a microtiter plate and printed onto SuperAmine2 glass slides (ArrayIt, Sunnyvale, CA), with a spot spacing of 250 μm. For ease of comparison, the highest
concentration of each dye sample present in the arrays is designated as 100% relative concentration.

Dye arrays were printed with a SpotBot II personal microarrayer robot (ArrayIt), using custom methodology to generate the XML code needed to drive the instrument (i.e., a SPOCLE file for the SpotBot II). First, the array was designed, using Photoshop to create an 8-bit RGB TIFF file. This image file is considered the map for the desired phantom, with each RGB color code representing a given sample (i.e., 3 “digits” of 0-255). A typical array image encompasses 20 x 20 virtual “pixels,” where each “pixel” corresponds to single-color square in the array of multiple squares in a Photoshop image area, and may contain multiple sub-arrays within this area. Specifically, the two array phantoms examined in this paper consist of six replicate subarrays (with 6 x 9 pixels dimensions). Second, a text file was created that acts as a “key” for the map. This text file indicates the microtiter plate well corresponding to a given color in the image. For example, the color red (255, 0, 0) may correspond to well A1. Additionally, a blank color (e.g., white) may be used to indicate that no printing should occur at a given location. Third, a MATLAB (Mathworks, Natick, MA) script was used to create the final XML file by linking the image file and the text file key. Additional inputs into the MATLAB script can be used to alter the spot spacing, the number and frequency of washing steps, whether to preprint spots in a separate location, and whether to repeat the arrays on a single slide or on multiple slides. The general work flow is illustrated in Fig. 1, where the printing was performed on glossy photo paper just to facilitate viewing, where the actual samples for analysis were prepared by printing spots on glass slides. An example of the MATLAB code is provided as supplementary material.

![Fig. 1.](image)

(a) A printing design is chosen or created in a program such as Photoshop. Here the U.S. Department of Commerce logo, containing 5 colors, is shown in a reduced image size and consists of five colors. (b) The colors represent printing wells (four of which correspond to four distinct colors, with white indicating that a spot should not be printed at that location). (c) The image and well codes are fed into MATLAB, which generates the appropriate code to drive the SpotBot II (see Media 1), as shown here. (d) The image was printed on Epson glossy photo paper to illustrate the printing procedure by showing the spots macroscopically. (e) A close-up comparison of a region of the inputted image and its corresponding printed image.
The microarray phantoms were imaged using a custom-built hyperspectral microscope whose detailed description, including the spectral bandwidth and results of background correction methods, is published elsewhere [5,14]. In this experiment, an OL490 tunable spectral light engine (Gooch and Housego, Orlando, FL) was coupled to a Zeiss Axiovert 100 inverted microscope, resulting in a spectral bandwidth of approximately 10 nm. Images were collected by sweeping the source in 5 nm increments from 400 nm to 700 nm. The transmitted light image was detected by a Hamamatsu Orca-R2 CCD camera (Hamamatsu, Japan). A focusing motor was used to correct for chromatic aberration of the objective lens at each wavelength. The OL490 selects each spectral bandwidth by turning only micromirrors corresponding to each bandwidth in the on position. Even when all mirrors of the OL490 are in the off position, some stray light exists and will be detected by the CCD. Here we define the dark signal of the system to include the electronic CCD noise and the signal from this stray light. Due to the change in the focus of the system, the dark signal of the hyperspectral microscope was found to be spectrally-dependent (i.e., the contribution of this stray light to the dark signal depends on the objective z-position). Therefore, a dark signal image was collected after each frame of the signal image. In addition to imaging the phantom sample, a reference image was collected from a region of the glass slide lacking spots. This area served as a reference to characterize the intensity and spatial homogeneity of the illumination light at each wavelength.

Post-processing of the signal and background data cubes was performed to transform the data set into absorbance values. A pixel-by-pixel correction method similar to that used by Geladi et al. [15] was performed to remove dark signal and nonhomogeneous illumination artifacts. In our analysis, the sample intensity detected at pixel x,y for a given wavelength, \( \lambda \), is written as \( I_{x,y,\lambda} \). A portion of this intensity is due to the dark signal inherent from the illumination source and camera, which we write as \( I_{\text{dark},x,y,\lambda} \). Similar intensities are determined for the background (i.e., the intensity detected if no sample is present), which are designated \( I_{0,x,y,\lambda} \) and \( I_{0,\text{dark},x,y,\lambda} \). Therefore, the absorbance at pixel x,y,\( \lambda \) can be calculated by

\[
A_{x,y,\lambda} = -\log_{10} \frac{I_{x,y,\lambda}}{I_{0,x,y,\lambda}} - \frac{I_{\text{dark},x,y,\lambda}}{I_{0,\text{dark},x,y,\lambda}}. \tag{1}
\]

![Sample Absorbance](image)

Fig. 2. Examples of detected images from the hyperspectral data cubes of a printed spot on glass. Signal and dark signal intensities are shown for the glass background datacube (\( I_0 \)), and the sample datacube (\( I \)). These intensities are used to calculate the absorbance image \( A \). All images are from the same spatial location on the CCD (i.e., the same x,y coordinates). Ring patterns within the hemispherical PEG-based spot are caused by scattering and refraction of the illumination light.
Examples of the acquired intensities and calculated absorbances for selected wavelengths are presented in Fig. 2.

3. Results and discussion

A microarray pattern consisting of spots containing individual or mixtures of three dyes (AR, BBR, and EG) was designed and analyzed (Fig. 3). The array contains 5%–100% of each individual dye as well as mixtures of two or three dyes. Additional spots of PEG without any dye are also included and serve as reference signals. An absorbance image of the printed microarray at 695 nm is shown in Fig. 3(b). The spots are hemispherical, consisting of PEG and dissolved dyes. In general the spots were uniformly printed; however some changes in the spot size occurred, likely caused by a minor obstruction in the tip of the SpotBot during printing. The imaged spots exhibit wavelength-dependent absorption and scattering signals. For example, the close-up view of a PEG spot can be observed in Fig. 2. The detected absorbance at a pixel depends strongly on the scattering and refraction of light by the hemispherical spot. Such factors complicate quantitative hyperspectral analysis, but are representative of some of the challenges that exist in biomedical samples. Because the demonstration of the preparation of these hyperspectral imaging phantoms is the primary focus of this paper, analysis of the microarrays here was performed by averaging the signals from individual spots. First, the spots were identified by thresholding the image at the 695 nm frame. The mean absorbances for all pixels within a spot were determined for each wavelength. Additionally, the mean background intensity was determined from all pixels outside the spots. Sample signals were corrected using this mean background signal.

The mean scattering signals from within a printed spot can be observed by monitoring the recorded absorbance from the reference spots containing only PEG (Fig. 3(c), top). An individual dye spectrum is the combination of the dye absorbance profile and the scattering signal. Small spectral differences for the PEG (scattering) spots were observed across the field of view as a result of changes in the scattering due to spot shape or position. As detailed analysis of the scattering signal is outside the scope of this paper, approximate methods were performed to simply remove the scattering components. For example, to recover a reference “pure” dye signal, the spectra of the 100% dyes were obtained by subtraction of the most similar PEG signals.

Fig. 3. (a) A pattern designed in Photoshop (shown with identifying notations), was used to generate a test microarray that consists of PEG spots containing varying amounts of three spectrally dissimilar dyes (AR, BBR, EG). (b) This sample was imaged by a hyperspectral microscope (shown here for 695 nm illumination). Scalebar is 200 μm. (c) The absorption spectra of dye spots measured by the hyperspectral microscope are shown before (top) and after (bottom) subtraction of the PEG background.
A similar array phantom was made to compare two spectrally similar dyes, AR and NC. Single dye concentrations were varied from 1% to 100%, and a large array of mixtures were printed (not shown). Examples of the 100% relative concentration spectra, prior to subtraction of the PEG signal, are shown in Fig. 4. This pair of dyes provides a test platform for analysis methods that may seek to differentiate two biological chromophores having extensive overlap (e.g., oxygenated and deoxygenated hemoglobin).

More intensive hyperspectral imaging analysis of such phantoms has been presented elsewhere, comparing the Least Absolute Shrinkage and Selection Operator (LASSO) and spatial LASSO [16]. However, a simplified analysis was performed in the current study by least-squared fitting of a linear combination of the dye spectra. The spectra obtained from the 100% dye spots were used to fit the remaining microarray spots (including the PEG spots), using a linear combination of the dye signals, and including a constraint of non-negativity. Again, the best-matched PEG signal was used in this fitting to match the scattering signal. For example, for the two dye phantom the fitting function was

\[
A_{\text{fit},\lambda} = c_{AR}A_{AR100,\lambda} + c_{NC}A_{NC100,\lambda} + A_{PEG,\lambda} + \text{offset},
\]

where \(A_{\text{fit},\lambda}\) is the fitted absorbance at wavelength \(\lambda\), \(c_{AR}\) and \(c_{NC}\) are the relative concentrations of dyes AR and NC, \(A_{AR100,\lambda}\) and \(A_{NC100,\lambda}\) are the PEG subtracted absorbances of the dyes at 100% relative concentration, and \(A_{PEG,\lambda}\) is the absorbance of the best fit PEG spectrum. \(A_{PEG,\lambda}\) is calculated directly from Eq. (1), while \(A_{AR100,\lambda}\) and \(A_{NC100,\lambda}\) are calculated by correcting the signal calculated in Eq. (1) by the appropriate PEG signal (resulting in PEG corrected spectra similar to the lower spectra in Fig. 3(c)). Figure 5(a) displays selected, experimentally measured and fitted dye spectra from the microarray appearing in Fig. 3. Figure 5(b) shows example fittings from the two spectrally similar dyes. The experimental and fitted spectra match very well in these cases. Deviation is noticeable near the extreme wavelengths (i.e., 400 nm and 700 nm) where the illumination intensity is weaker, which results in larger errors in the absorbance calculation. The linear mixing method worked reasonably when only one dye...
was present within a spot, even when significant spectral overlap existed between the two candidate dyes (see Table 1). In these cases, the measured and fitted results were very consistent across replicate microarrays printed on the same slide. Similar results were observed for the phantom with the three dyes. However, spots that contained dye mixtures were more challenging, and the results were often erroneous. For example, the fitting often found more NC present than the real value when this dye was mixed with AR (see Table 1). While, the linear mixing method often was superior for the single AR or NC dyes, analysis of the complex dye mixtures was better performed by the spatial LASSO, which typically resulted in dye concentrations that more closely matched the actual value.

Table 1. Expected and calculated concentrations from linear mixing analysis of two dyes: AR and NC. Concentrations are relative to the maximum printed solution concentration. Standard deviation is calculated from the analysis of three different microarrays on the same slide.

| Sample          | Expected [AR] | Calculated [AR] | Expected [NC] | Calculated [NC] |
|-----------------|---------------|-----------------|---------------|-----------------|
| AR050:NC100     | 50            | 0.16 ± 0.09     | 100           | 1.40 ± 0.170    |
| AR050:NC050     | 50            | 0.25 ± 0.220    | 50            | 0.46 ± 0.266    |
| AR010:NC020     | 10            | 0.06 ± 0.055    | 20            | 0.23 ± 0.148    |
4. Conclusion

Spectral analysis of biological systems generally is quite difficult, and care must be taken to properly test the measurement process. The advantage of using a microarray phantom is that it can be tailored to a given biological target. For example, remote sensing analysis of phytoplankton pigments is complicated due to the relationships between pigment concentrations [17]. In performing in vivo melanoma diagnostics, the spectral differentiation between malignant melanoma and dysplastic nevi is crucial [18]. For both of these biological phenomena, initial testing of the statistical methods for image analysis could be performed by the preparation, printing, and analysis of microarray phantoms containing interacting dyes or pigments. The modeling of scattering systems is particularly relevant to the hyperspectral imaging of agricultural products, which require the decoupling of significant scattering and contour contributions to the measured spectra [19]. Similar issues also arise when analyzing interacting cell populations, such as those comprising complex biofilms [20]. A microarray platform with spots of varying size could be used as a test platform to deduce the influence of spot curvature on the analysis results.

Small scale systems, such as those illustrated here to print custom-designed microarray phantoms are attractive tools. In particular, they provide the ability to manufacture a large number of standards, which can be readily shared across institutions and used to compare different hyperspectral imagers and analysis methods. Additionally, the preparation of a microarray allows quick exploration of the relevance and relative quantities of various components considered for tissue phantoms. The spatial sizes can be scaled up for larger area standards once the desired composition is determined from the microarray analysis.

Appendix A: Supplementary material

Example MATLAB code to program the SpotBot II is provided as supplementary material (Media 1).

Appendix B: Disclaimer

Certain commercial equipment, instruments, or materials are identified in this manuscript are to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Acknowledgments

This research was supported by the NIST Innovative Measurement Science Program on optical medical imaging. Dr. Ralph Nossal was supported by intramural funds of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH.