Title
A standardized data specification for reproducible, verifiable structured light and multispectral image sets

Journal
DESIGN AND QUALITY FOR BIOMEDICAL TECHNOLOGIES XIX, 13840

ISSN
1605-7422

ISBN
978-1-5106-9593-1

Authors
Campbell, Christopher A
Nudurupati, Pratik
Durkin, Anthony J

Publication Date
2026

DOI
10.1117/12.3081186

A standardized data specification for reproducible, verifiable structured light and multispectral image sets

Christopher A. Campbell*[a], Pratik Nudurupati [a], Anthony J. Durkin[a,b]

[a]Beckman Laser Institute and Medical Clinic, University of California Irvine, 1002 Health Sciences Road East, Irvine, CA 92697; [b]Department of Biomedical Engineering, University of California, Irvine

ABSTRACT

Spatial Frequency Domain Imaging (SFDI) is a quantitative wide-field imaging technique that uses light and diffuse optical spectroscopic principles to rapidly generate tissue optical property images. However, SFDI datasets present significant workflow challenges: they can be large (on the order of gigabytes), contain multiple image formats, and require computationally intensive processing. Additionally, SFDI images typically require calibration measurements, further complicating data organization. The proliferation of calibration data, imaging data, and processing options creates technical overhead and leads to reproducibility issues.

While multiple "Open SFDI" initiatives exist, no standardized storage format has emerged, making the replication of published SFDI research difficult. Commercial solutions like Modulim's directory-based format contain trade secrets and lack open-source processing options.

We are developing an open standard for SFDI images, extensible to multispectral images, dubbed SpectralMaps. This specification leverages modern Xarray data structures saved in Zarr format, offering fast loading of complete datasets or selected data chunks. Zarr permits lossless compression with partial loading of data, making data loading faster and partial data processing possible without loading the entire dataset. Metadata for measurements and processing history are included in the file. Datasets are identified by hashes included for verification. This specification provides a unified, transparent framework for storing, sharing, and processing SFDI data, reducing technical overhead and improving reproducibility across research groups. By adopting an open, extensible standard, the community can accelerate method development, enable cross-platform interoperability, and foster collaborative validation of SFDI techniques.

Keywords: Spatial Frequency Domain Imaging, multispectral imaging, data processing, standardization, open-source

1. INTRODUCTION

Spatial Frequency Domain Imaging (SFDI) is a non-contact, wide-field imaging modality pioneered at the Beckman Laser Institute that uses light and diffuse optical spectroscopic principles to generate quantitative maps of tissue optical properties. Projecting structured illumination on a tissue target—typically in the form of sinusoidal light patterns—at multiple wavelengths and spatial frequencies (see Figure 1), the system employs light propagation models to decouple the effects of absorption from reduced scattering. This approach allows for the interrogation of subsurface tissue volumes up to 5 mm in depth, enabling the visualization of functional constituents like oxy- and deoxyhemoglobin, water, and lipids, while simultaneously providing structural insights into cellular and extracellular components such as collagen.

SFDI has demonstrated significant utility across a broad spectrum of clinical applications, including burn severity assessment, predicting skin flap survival during surgical reconstruction, monitoring cerebral hemodynamics, and skin cancer detection.[1–4] Beyond the medical field, the technique's depth-sensitive mapping capabilities have proven effective for industrial and agricultural purposes, such as identifying subsurface bruising in produce and evaluating the quality of other perishable foods. Although these datasets present unique workflow challenges due to their large size and calibration requirements, SFDI remains a powerful, non-invasive tool for objective tissue characterization in both research and clinical settings.[5,6]

*cacampbe@uci.edu

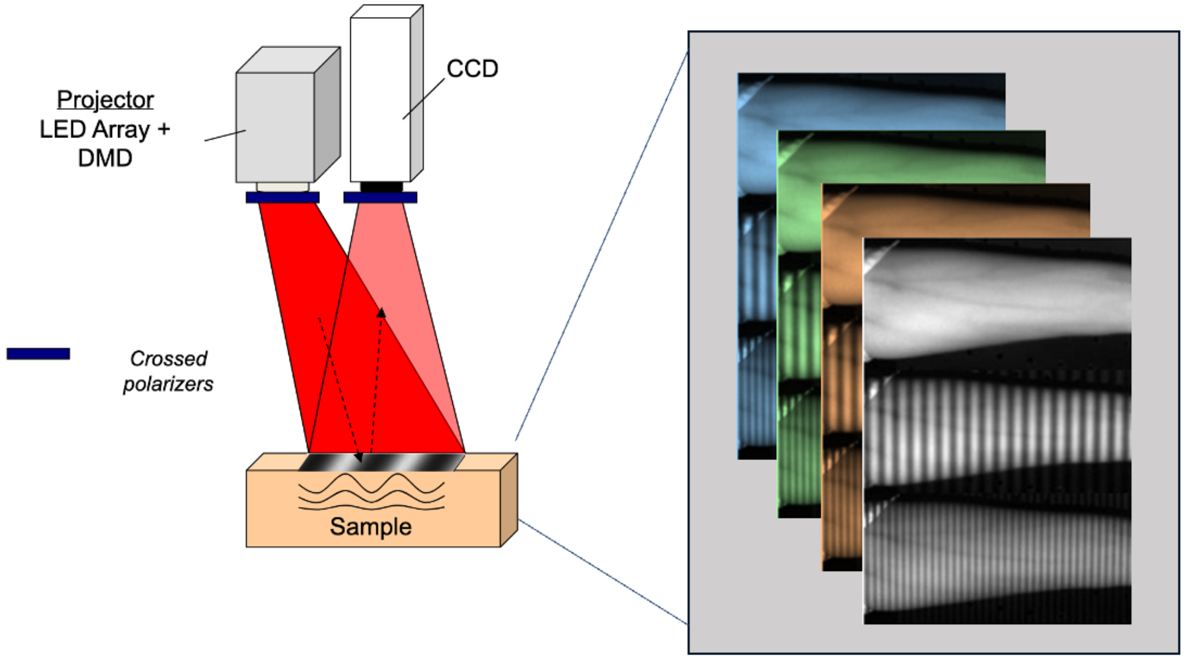

Figure 1. SFDI Principle of operation. In a typical configuration, a projector illuminates the region of interest (ROI) using a series of sinusoidal patterns, with the phase shifted sequentially by 120 degrees. This is repeated for each wavelength. A camera records the reflected light. For quantitative measurements of tissue reflectance, this measurement is duplicated on a tissue simulating phantom having known optical properties, which can be used to calibrate the tissue measurement.

One of the aspects of SFDI that limits more widespread usage as a research tool is that SFDI datasets present complicated workflow requirements. They are large (on the order of gigabytes), may contain multiple image formats, and have unique, computationally intensive processing needs. Fig. 1(a) shows a partial view of an example dataset. As is typical of biomedical imaging techniques, SFDI images often require a second measurement for calibration, of a specific optical phantom with known properties. Finally, acquisition settings and processing options must be recorded somewhere.

The proliferation of calibration data, imaging data, and acquisition and processing options can incur significant technical overhead and easily lead to errors when it comes to reproducibility. In this work, we present our plan for ongoing development of a standardized, open, extensible format specification for SFDI data sets, generalizable to multispectral imaging, using open-source tooling.

2. CURRENT LANDSCAPE

With NIH-led initiatives to increase the open nature of science produced in the US, multiple projects have been produced using the name "Open SFDI." The first we are aware of can be found at opensfdi.org[7]. A pre-compiled program is provided for processing a TIF stack, along with a tutorial on how to go about writing software to process data. More recently, a project out of Brazil was reported on, combining low-cost hardware and some simple machine learning models.[8] The latter, at least, lowers the bar for entry into SFDI research. However, there are several published methods for producing and processing structured illumination data in the literature, with no easy means of replication.[9–13]

In seeking solutions to this problem, one of the examples that we can look to in terms of a paradigm for addressing these kinds of issues is in the hyperspectral imaging (HSI) literature. HSI, which is a common tool used in satellite imagery, has a similar storage problem to SFDI. This area has seen more interest and development, resulting in numerous competing specifications. Of note, these specifications appear to frequently use the netCDF format.[14] SFDI data has specific needs in addition to the needs of typical hyperspectral data. The proposed specification will prefer the more modern Xarray data structure, saved to disk in the Zarr format.[15,16]

3. DESIGN GOALS

Minimize computational overhead

The SpectralMaps format prioritizes speed and efficiency in handling data. To facilitate this goal, it should be possible to load all or part of a dataset from disk, and to process all or part of the data.

Fast retrieval

Store and access files efficiently, either locally or over the network. This includes the ability to open a subset of images, for example a square region of interest over a specific wavelength range.

Sufficient metadata included for verification and repeatability

One should not need to rely on file naming conventions to understand the provenance of the data in the SpectralMaps file. Coordinates like wavelengths, spatial frequencies, exposure times, and equipment used should be included within. If the dataset has been processed, include the information necessary to reconstruct the processing workflow from raw datasets. Extraneous metadata, for example referring to acquisition options that don't exist, should be omitted.

Include a means for verifying data integrity

To reasonably confirm that the file has not been corrupted, include a list of hashes that should match portions of the dataset.

To meet these goals, we selected the Xarray format, with support for saving to disk as a Zarr DataStore (default), with NetCDF as an alternative option.

4. WORKING WITH SPECTRALMAPS DATA

Built on Xarray DataTrees

```
/ (root)
├── meta
│   ├── instrument (family, name, software version, ...)
│   ├── measurement (study, subject, timestamp, …)
│   └── attributes (data hashes)
├── main
│   ├── sfdi
│   ├── dark
│   ├── planar
│   ├── profile
│   └── alignment
└── etc
    ├── temperature
    ├── humidity
    └── RGB
```

Figure 2. A typical structure containing SFDI and related data. The root node consists of multiple branches grouped according to the format of the data saved within. Structured illumination images and other data collected with the primary camera is stored in the "main" node. Metadata associated with the measurement are kept in "meta." Additional data may be stored under "etc."

Data are saved in the Xarray format[15], which permits various ways of indexing and slicing data using increments or coordinates. Each contiguous data cube is saved as an Xarray DataArray, and multiple arrays can make up a branch in the overall structure, which is a DataTree. For SpectralMaps data, we currently specify three "branches" by default: meta, containing information about the measurement collection and processing parameters; main, containing images collected through the main camera; and etc for data with a different pixel counts than data collected through the main camera. Examples include single-point thermocouple readings or an associated RGB image with a resolution distinct from the main camera. The file organization is demonstrated in Figure 2.

Branches and leaves in the tree can be accessed via a variety of methods, the simplest of which is decimal-delimited, e.g, `sfdata=datatree.structured.sfdi.data`. Alternatively, a dictionary-like syntax can be used, as in `sfdata=datatree["structured"]["sfdi"].data`. The selected multidimensional array can be further sliced, for example to pull a single rectangular ROI from the set, using any of the indexing methods XArray permits.

5. COMPARISON WITH EXISTING FORMATS

Table 1. Comparison of SFDI data storage formats.

	Modulim	**TIFF Stack**	**SpectralMaps**
Number of files saved	Many	One, Two, or Many	One
File types	Binary and .xml	.tif and .xml	.zarr or .hdf5
storage	Uncompressed	Uncompressed	Default compressed/chunked
metadata	File name and .xml	Within file, optional .xml	Entirely within file
standardization	Yes, but opaque	None	Yes, open source
Processing	Proprietary	Specific to research group	Open source encouraged
checksums	No	No	Included in file

Table 1 tabulates the differences between commonly used formats and SpectralMaps. Key aspects include the compressed, chunked storage on disk of SpectralMaps data, meaning SpectralMaps data takes up less disk space by default while allowing part or all of the data to be loaded.

6. CONCLUSIONS & FUTURE WORK

The proposed open data specification addresses the fundamental bottleneck of reproducibility in SFDI research by providing a standardized, portable framework for high-dimensional SFDI and multispectral datasets. By leveraging the Xarray DataTree structure and Zarr Store format, this specification enables the efficient handling of gigabyte-scale data through lossless compression and partial data loading, which are essential for both local inspection and processing, as well as larger-scale processing operations. This move toward standardization is a vital step in transitioning SFDI from a specialized research technique to a robust, verifiable clinical tool, facilitating the open data exchange necessary for rigorous, multi-center scientific validation.

Future efforts, as funding is secured, will prioritize the development of open-source conversion utilities to migrate existing, opaque commercial directory structures into this standardized format, thereby lowering the barrier for community adoption. We intend to further integrate this specification with high-performance computing pipelines, such as Nextflow, to support large-scale longitudinal clinical trials and the real-time processing requirements of modern surgical environments. We anticipate this specification will readily accommodate emerging high-speed hardware configurations, such as multi-tap CMOS image sensors, which require specialized handling of simultaneous multi-wavelength and ambient light suppression data. Ultimately, by fostering a community-driven design process, we hope to establish this specification as an extensible standard for diverse biophotonic applications, ranging from the early detection of nascent cutaneous neurofibromas to the quantitative monitoring of Thyroid Eye Disease.

ACKNOWLEDGEMENTS

The authors thankfully recognize the support provided by the NIH, including the National Institute of General Medical Sciences (NIGMS) (R01GM108634) and NIAMS (5T32AR080622). Furthermore, this material is based, in part, upon technology development supported by the U.S. Air Force Office of Scientific Research (FA9550-23-1-0685). Moreover, we express our gratitude to the Arnold and Mabel Beckman Foundation. The content is the sole responsibility of the authors. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect or represent the official views of the NIGMS, NIH, U.S. Air Force, or Department of Defense.

REFERENCES

1. D. M. Burmeister et al., "Utility of spatial frequency domain imaging (SFDI) and laser speckle imaging (LSI) to non-invasively diagnose burn depth in a porcine model," Burns **41**(6), 1242–1252 (2015) [doi:10.1016/j.burns.2015.03.001].
2. A. Ponticorvo et al., "Quantitative assessment of graded burn wounds in a porcine model using spatial frequency domain imaging (SFDI) and laser speckle imaging (LSI)," Biomed. Opt. Express **5**(10), 3467–3481 (2014).
3. C. Crouzet et al., "Intrinsic, widefield optical imaging of hemodynamics in rodent models of Alzheimer's disease and neurological injury," Neurophotonics **10**(2), 020601 (2023) [doi:10.1117/1.NPh.10.2.020601].
4. R. B. Saager et al., "A Light Emitting Diode (LED) Based Spatial Frequency Domain Imaging System for Optimization of Photodynamic Therapy of Nonmelanoma Skin Cancer: Quantitative Reflectance Imaging," Lasers Surg. Med. **45**(4), 207–215 (2013) [doi:10.1002/lsm.22139].
5. X. He et al., "Spatial frequency domain imaging for detecting bruises of pears," J. Food Meas. Charact. **12**(2), 1266–1273 (2018) [doi:10.1007/s11694-018-9740-5].
6. Z. Sun et al., "Development of a multispectral spatial-frequency domain imaging system for property and quality assessment of fruits and vegetables," Comput. Electron. Agric. **214**, 108251 (2023) [doi:10.1016/j.compag.2023.108251].
7. M. B. Applegate et al., "OpenSFDI: an open-source guide for constructing a spatial frequency domain imaging system," J. Biomed. Opt. **25**(1), 016002 (2020) [doi:10.1117/1.JBO.25.1.016002].
8. M. S. Sampaio et al., "Portable spatial frequency Domain imaging using Open-source tools," Instrum. Sci. Technol., 1–20, Informa UK Limited (2025) [doi:10.1080/10739149.2025.2513279].
9. D. J. Cuccia et al., "Quantitation and mapping of tissue optical properties using modulated imaging," J. Biomed. Opt. **14**(2), 024012 (2009) [doi:10.1117/1.3088140].
10. S. Gioux et al., "Three-dimensional surface profile intensity correction for spatially modulated imaging," J Biomed Opt **14**(3) (2009) [doi:10.1117/1.3156840].
11. E. Aguénounon et al., "Real-time optical properties and oxygenation imaging using custom parallel processing in the spatial frequency domain," Biomed. Opt. Express **10**(8), 3916–3928, Optica Publishing Group (2019) [doi:10.1364/BOE.10.003916].
12. M. Ghijsen et al., "Real-time simultaneous single snapshot of optical properties and blood flow using coherent spatial frequency domain imaging (cSFDI)," Biomed. Opt. Express **7**(3), 870–882 (2016).
13. L. Belcastro, H. Jonasson, and R. B. Saager, "1 Multi-frequency SFDI: a depth-resolved optical scattering model 2 to isolate optical properties in thin layers of skin."
14. R. Rew et al., "Unidata NetCDF," UCAR/NCAR - Unidata (1989) [doi:10.5065/D6H70CW6].
15. S. Hoyer and J. Hamman, "xarray: N-D labeled Arrays and Datasets in Python," J. Open Res. Softw. **5**(1) (2017) [doi:10.5334/jors.148].
16. "Zarr," Zarr, <https://zarr.dev/> (accessed 9 July 2025).