SAND2021-5403C

A Process to Colorize and Assess Visualizations of Noisy X-Ray Computed Tomography Hyperspectral Data of Materials with Similar Spectral Signatures

Joshua Clifford, Emily Kemp, Ben Limpanukorn, Edward S. Jimenez, *Sandia National Laboratories*

***Abstract*— Dimension reduction techniques have frequently been used to summarize information from high dimensional hyperspectral data, usually done in effort to classify or visualize the materials contained in the hyperspectral image. The main challenge in applying these techniques to Hyperspectral Computed Tomography (HCT) data is that if the materials in the field of view are of similar composition then it can be difficult for a visualization of the hyperspectral image to differentiate between the materials. We propose a novel method of preprocessing and summarizing hyperspectral data in a single colorized image and novel measures to assess desired qualities in the resultant colored image, such as the contrast between different materials and the consistency of color within the same object. Proposed processes in this work include a new majority-voting method for multi-level thresholding, binary erosion, PAM clustering for grouping pixels into objects (of homogeneous materials), smoothing splines along the spectral dimension to estimate the underlying signature, dimension reduction with UMAP to assign colors, and quantitative coloring assessment with developed measures. This process produced a colorization with higher within-object smoothness and between-object contrast than other methods tested. These results have the potential to create more robust material identification methods from HCT data that has wide applications in industrial, medical, and security-based applications for detection and quantification, including visualization methods to assist with rapid human interpretability of these complex hyperspectral signatures.**

Keywords—x-ray computed tomography, hyperspectral imaging, colorization, dimension reduction, clustering, smoothing splines

I. Introduction

This paper explores several preprocessing and dimension reduction methods with the goal of representing all channels in a hyperspectral computed tomography (HCT) dataset with one colorized image, portraying objects with maximum smoothness within the objects (each is a homogeneous material) and contrast between objects (to distinguish varying concentrations) to assist with human interpretability. Past research has visualized hyperspectral data using preprocessing and dimension reduction techniques. For example, [1] used principal component analysis (PCA), self-organizing maps, and t-distributed stochastic neighbor embedding (t-SNE) to reconstruct mass spectrometry imaging data into a colorized visualization. Similarly, in [2], PCA and a multi-class support vector machine were used to classify regions within images, and a colorized image was produced as a by-product. Smoothing splines were used in [3] for noise removal and peak detection in hyperspectral imaging. Unlike previous work which dealt with hyperspectral data of materials with distinct spectral signatures, our work focuses on the challenge of visualizing materials that have relatively similar material composition and concentrations.

II. Methods

A. Data

The HCT dataset used to develop the methods was generated by PHITS [4], a general purpose Monte Carlo particle transport simulation code. The dataset contains nine cylindrical objects placed in a grid pattern and are imaged in simulation with a patterned anode that contains tungsten, molybdenum, and silver to create a distinct hyperspectral x-ray spectrum from a 225kVp electron beam and detected by an ideal hyperspectral x-ray detector that channelizes the detected photons into 128 energy channels spread out across 300keV [5, 6]. Each object is a different concentration of a liquid mixture of two materials.

B. Preprocessing Techniques

Our approach uses a series of preprocessing methods that isolate objects and spectrally and spatially de-noises the data on a per-object basis. First, negative intensity values in the hyperspectral data were replaced with zeros since the negative values were an artifact of the reconstruction method used to generate the synthetic data. The next stage of the pipeline was to generate an image mask that isolates the objects from the background. A preliminary mask was first created with a "majority vote" thresholding algorithm that performs multi-level thresholding with the Multi-Otsu algorithm on each channel of the image and then combines the per-channel masks into a single mask. For a pixel in the final mask to be of a certain class, the majority of the corresponding pixels in that same location in the per-channel masks must be of that class. This thresholding method avoids the need for manual selection of a single specific channel to use for multi-level thresholding. This preliminary mask is further refined by applying binary erosion to remove any remaining non-closed artifacts and object borders. Once the objects were isolated by the image mask, we explored several methods to achieve consistency of colors within homogeneous materials. Spatial smoothing was done by convolving a 2D box filter or a 2D Gaussian filter over the hyperspectral data for each channel. Smoothing was also done along the spectral dimension by fitting a cubic smoothing spline across the 128 channels. Another method to improve color consistency within an object was to use the partitioning around medoids (PAM) algorithm to group the pixels of each object together and then estimate the object's spectral signature with a smoothing spline fit to all pixels in an object. The pixels were clustered based on the intensity per channel augmented with the location of the pixel in the image to encourage clustering based on spatial locality.

C. Dimension Reduction for Coloring

While numerous methods were tested, the methods presented in this paper's results include t-SNE and Uniform

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525

Manifold Approximation and Projection (UMAP), both widely applied to visualize high-dimensional data. To colorize the objects, the 128-channel hyperspectral data were transformed into a three dimensional space using t-SNE or UMAP, and each dimension was [0, 255] normalized for use as a RGB color value.

D. *Colorization Assessment Metrics*

There are several specific attributes we wish to measure in the colorized image: the variation/inconsistency within each object, the perceptual contrast between the objects, and the average box filter color variation throughout the object (how speckly the object appears). We created several metrics to estimate each of these specific attributes. All three metrics are based on the notion that humans perceive the RGB color model as an additive model. We can then formulate the variance of a variable, Y', which represents the human perception of the red (R), green (G), and blue (B) channels, as:

$$\mathrm{Var}(Y') = \mathrm{Var}(R)+\mathrm{Var}(G)+\mathrm{Var}(B)+2\mathrm{Cov}(R,G)+2\mathrm{Cov}(R,B)+2\mathrm{Cov}(G,B) \quad (1)$$

Each of the three developed metrics (inconsistency, contrast, speckliness), utilize this formulation to measure each of their respective attributes. Each metric has the relationship that the larger it is, the more of that respective attribute is present in the colorized image.

III. RESULTS

Below are example UMAP colorized images for each step of the preprocessing pipeline to remove pixels not in an object.

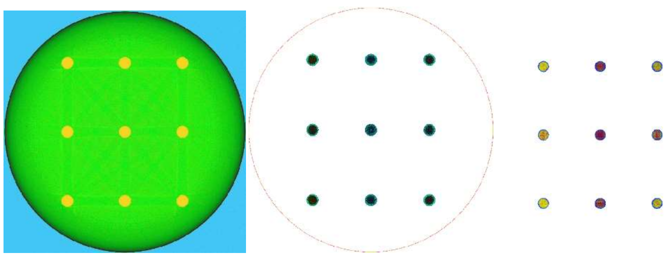
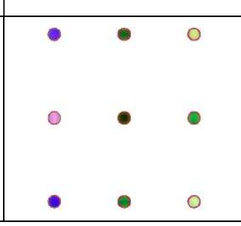
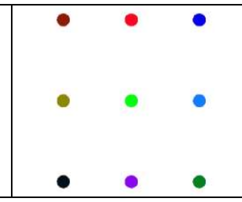


Fig. 1. Illustrations of preprocessing steps. From left to right: no preprocessing (raw data), majority-voting thresholding, and thresholding plus erosion.

Below are example colorized images for alternative paths tested for preprocessing images to color (spatial smoothing versus object clustering and smoothing splines). Metric calculation results, including inconsistency (I), contrast (C), and speckliness (S), are also included for each, along with those for thresholding and erosion only.

TABLE I. EXAMPLE COLORIZATION IMAGES AND METRICS

Preprocessing	**UMAP Image**	**Metrics**
Thresholding + Erosion	*(See Figure 1)*	I = 0.0009 C = 0.1598 S = 0.0409
Spatial Smoothing Filter (Gaussian, s = 1)		I = 0.0050 C = 0.2000 S = 0.0233
Clustering + Object-Level Smoothing Splines		I = 0.0000 C = 0.5069 S = 0.0000

IV. DISCUSSION/CONCLUSION

The metrics provided quantitative validation of the qualitative assessment of each object. The majority-voting thresholding in combination with erosion were extremely effective at removing pixels not associated with objects to focus coloring on the objects of interest, a necessary step to avoid noise in the colorization due to photon scattering and other artifacts as well as allow contrast to begin to emerge between the objects. Consistency of within-object coloring was still difficult to achieve. The spatial smoothing filter helped coloring centers of objects but emphasized edge artifacts, which are likely due to beam hardening. Clustering each object and using splines to estimate the object's underlying hyperspectral function over the channels provided consistency within objects as well as, using UMAP, contrast between objects for the final colorization scheme. The underlying function estimation is likely biased by the edge artifacts, so ongoing research will investigate proper corrections for this phenomenon, which may also make spatial smoothing filters a viable alternative to clustering and splines for preprocessing before colorization. In addition, further research will be done into the robustness and interpretability of t-SNE, UMAP, and other colorization methods as well as their consistency between images with different types of materials (including non-liquids) and concentrations of mixtures.

ACKNOWLEDGEMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] J. M. Fonville, et al., “Hyperspectral Visualization of Mass Spectrometry Imaging Data,” *Anal. Chem.*, vol. 85, no. 3, pp. 1415–1423, 2013.

[2] S. Ranjan, D. R. Nayak, K. S. Kumar, R. Dash, and B. Majhi, "Hyperspectral image classification: A k-means clustering based approach," *4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2017, pp. 1-7.

[3] A. Bjorgan, and L. L. Randeberg, “Application of smoothing splines for spectroscopic analysis in hyperspectral images,” *SPIE BiOS*, 2019.

[4] T. Sato, et al., “Features of Particle and Heavy Ion Transport code System (PHITS) version 3.02,” *J. Nucl. Sci. Technol.*, vol 55, pp. 684-690, 2018.

[5] G. Dalton, B. Limpanukorn, E. Kemp, J. Clifford, and E. S. Jimenez, “Monte-Carlo modelling and design for a high-resolution hyperspectral computed tomography system with a multi-material patterned anode for material identification applications,” *SPIE Optics+Photonics*, San Diego, CA, USA, Aug. 1-5, 2021. (Accepted)

[6] E. S. Jimenez, K. R. Thompson, A. Stohn, and R. N. Goodner, "Leveraging multi-channel x-ray detector technology to improve quality metrics for industrial and security applications," in *Proc. SPIE 10393: Radiation Detectors in Medicine, Industry, and National Security XVIII*, 2017, doi: 10.1117/12.2275850.