Material Characterization using a Compact Computed Tomography Imaging Spectrometer with Super-resolution Capability

Simon Amann[1⋆], Mazen Mel[2⋆], Pietro Zanuttigh[2], Tobias Haist[1], Markus Kamm[3], and Alexander Gatto[3]

[1] University of Stuttgart, Institut für Technische Optik, 70569 Stuttgart, Germany
[2] University of Padova, Department of Information Engineering, 35131 Padova, Italy
[3] Sony Europe B.V., Stuttgart Technology Center, 70327 Stuttgart, Germany

Abstract Computed Tomography Imaging Spectrometer (CTIS) systems are snapshot hyperspectral imaging devices capable of capturing dense spectra of static as well as dynamic scenes. A three-dimensional hyperspectral cube is smeared across the spatial dimension via Diffractive Optical Element (DOE) and projected across multiple angles forming a two-dimensional compressed sensor image. In this paper we demonstrate material characterization and classification capability of a compact CTIS system leveraging spectral signatures. Then we propose an approach to simultaneously reconstruct and segment into regions corresponding to different materials hyperspectral images with enhanced spatial resolution from CTIS sensor measurements.

Keywords CTIS, spectral reconstruction, super resolution, optical characterization

1 Introduction

Hyperspectral Imaging (HSI) plays an important role in the field of optical characterization of materials [1]. It allows, for example, to distin-

⋆ Authors contributed equally.

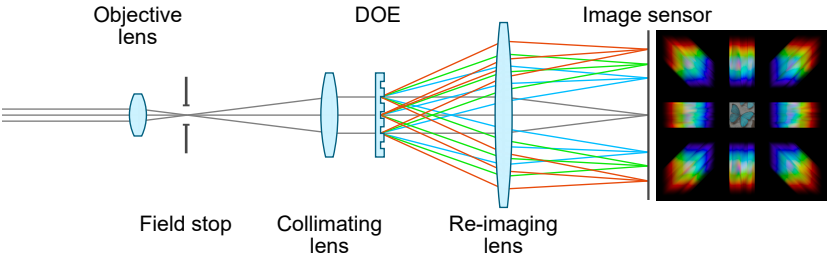


Figure 1: Optical layout of a commonly used CTIS system. Image based on [4].

guish or identify materials that look almost identical in a monochrome or color image. HSI-devices acquire a complete spectrum for each imaged object point. The resulting hyperspectral cube has three dimensions: the two spatial ones and the spectral dimension.

A Computed Tomography Imaging Spectrometer (CTIS) is based on a non-scanning (snapshot) technique [2]. Other methods in this area are the multi-aperture filtered camera and the pixel-level filter array camera [3]. They are both based on spectral filters. CTIS, on the other hand, uses a diffractive optical element (DOE) in combination with computational imaging algorithms. Figure 1 shows an optical layout of a commonly used CTIS system. The objective lens images the scene on the left to an intermediate image plane. There, it is cropped by a field stop, which defines the system's field of view. The collimating lens collimates the light, which is then spectrally dispersed by a diffractive optical element. A re-imaging lens creates the final sensor image. An example is shown on the right. It contains several higher diffraction orders arranged around the undiffracted zeroth order image of the scene. The higher diffraction orders are spectrally smeared. Blue light hits the sensor closer to the center than its red counterpart.

A reconstruction algorithm is needed to get the hyperspectral image from this spatio-spectral smeared sensor image. It solves a similar inverse problem as the reconstruction algorithms needed for computed tomography scanners. The different diffraction orders can be conceived of as two-dimensional projections of the three-dimensional hyperspectral-cube onto the image sensor. The Expectation-Maximization (EM) algorithm has been predominantly

used in CTIS image reconstruction [5]. The EM iteratively solve for the latent hyperspectral cube starting from an initial estimate. EM cannot handle priors and it is sensitive to the presumed noise and system model leading sometimes to poor reconstruction quality. Deep learning-based approaches have been devised to tackle the shortcomings of the EM solver: In [6] the authors used a sequential approach with a CNN followed by an EM solver wherein the CNN provides the initial estimate for the EM stage. Zimmermann *et al.* [7] proposed an end-to-end learning approach performing customized reshaping operations at the beginning to get an input shape suitable for 3D processing of high dimensional input data that is followed by a U-Net like architecture used to refine the estimated hyperspectral cube. We have recently proposed HSRN [8] tackling for the first time spectral reconstruction and spatial super-resolution from CTIS measurements. It allows to achieve a higher spatial resolution than that of the zeroth diffraction order while reconstructing accurate spectral information.

2 Method

We propose a two-stage approach for object classification using hyperspectral data captured by a CTIS system (see Figure 2). In the first stage we train our HSRN [8] architecture for hyperspectral reconstruction and spatial super resolution with up to $\times 5$ the resolution of the zeroth diffraction order for synthetic data. In the second stage, the reconstructed hyperspectral cubes are used to train a ResUnet [9] to perform semantic segmentation. The network produces two segmentation maps, one corresponding to object classes and the other underlining whether those objects are real or fake. Note that the two networks are trained separately. In more details, we use slightly modified architectures of both networks for better reconstruction quality and to avoid over-fitting. For HSRN [8] we increase the number of filters within the refinement network from 64 to 128 for all convolution layers and set the super-resolution factor to 5 for synthetic data and 2 for real data while keeping the rest of the architecture unchanged. For ResUnet [9] we use the modified architecture shown in Figure 2, the network has two output layers, one for each segmentation task. We train both networks for 500 epochs and use the training settings of HSRN suggested

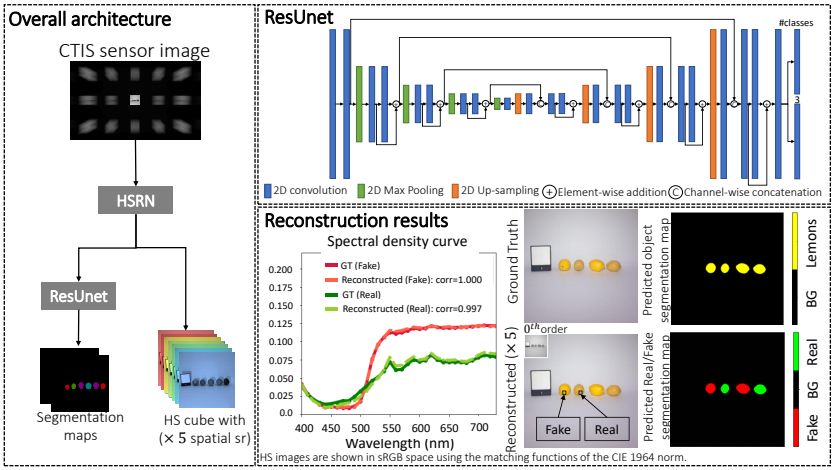


Figure 2: Left: Proposed two-stage architecture for hyperspectral image reconstruction and semantic segmentation, the two networks are trained separately. Upper right: The slightly modified ResUnet architecture used to learn object class and real/fake segmentation maps. Lower right: A reconstruction example with $\times 5$ spatial super-resolution and the corresponding segmentation maps, we also show spectral density curves of two selected image regions (real and fake lemons) along with the Pearson correlation coefficient to assess the accuracy of the reconstructed spectra.

in [8]. The cross-entropy loss is used to train the ResUnet.

3 Datasets

Synthetic data We use Fourier optics to simulate CTIS sensor images using hyperspectral cubes from FVgNet dataset [10] containing 252 labeled scenes of real and fake fruits and vegetables. A DOE that generates a structure with 5×3 diffraction orders is used in the simulation (see Figure 2). The simulated zeroth order has a spatial resolution of 102×102 pixels while the ground truth hyperspectral cubes have 510×510 pixels which corresponds to a $\times 5$ spatial super-resolution of the reconstructed cube. As in [10], we use a spectral range of $[400nm, 730nm]$ with 34 spectral bands. We chose randomly 80% of the scenes as training data and the rest for testing, random vertical and

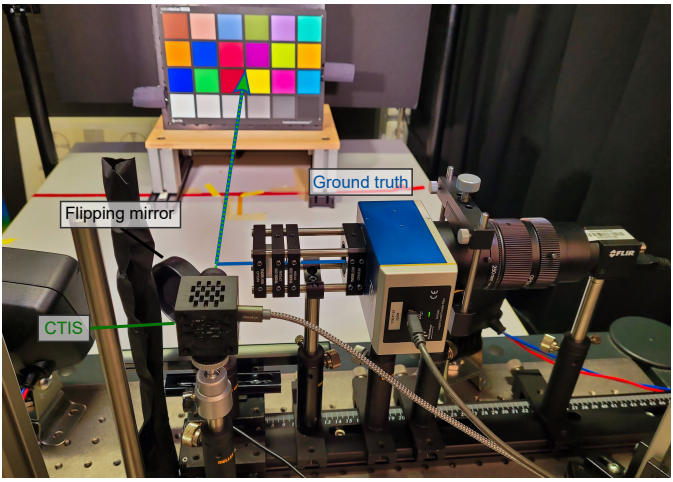
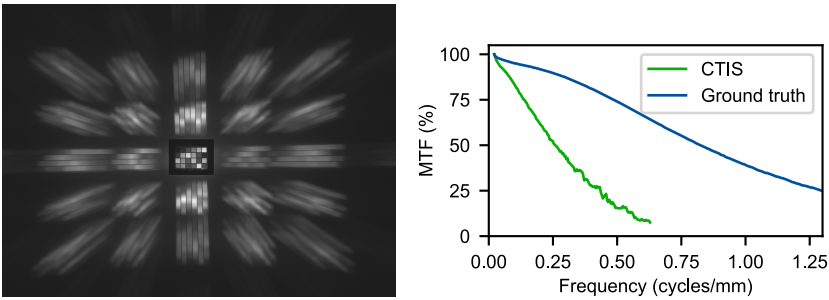


Figure 3: Photo of the miniaturized prototype together with the ground truth setup.

horizontal flipping is used as data augmentation.

Real data We have implemented a setup to validate that our reconstruction method also works on real world CTIS data. A photo of the system is shown in Figure 3. For the dataset needed to train our model, we always acquire a CTIS measurement together with a ground truth measurement. Our CTIS system is built with off-the-shelf lenses, a computer-generated hologram, a commercial smartphone lens and a 13 MP monochrome smartphone image sensor. The dimensions of the prototype are only 36.0 mm $\times$ 40.5 mm $\times$ 52.8 mm. This small size is achieved by using a Galilean instead of the commonly used Keplerian beam expander. Its diagonal field of view is 29°. The DOE creates a 5×5 arrangement of the diffraction orders. The zeroth order image size is 420×312 pixels, which corresponds to around 10% of the horizontal and vertical sensor size. Filters are used to limit the captured spectral range from 470 nm to 700 nm. Each CTIS measurement is made of two images captured with different exposure times (7.8 ms and 250 ms). This is needed to get an image with a well exposed zeroth order and one with well exposed higher diffraction orders. Our

prototype is therefore not a single-shot camera. Figure 4(a) shows a sample acquisition of a ColorChecker. The zeroth order part of the image taken with the longer exposure time is exchanged with that of the shorter exposure time. More information about a similar system can be found in [11]. Amann et al. [11] use the same prototype, just with a different shortpass filter.



(a) Sensor image **(b)** Modulation transfer function (sagittal)

Figure 4: Sensor image of the CTIS prototype and MTF measurement results comparing the CTIS prototype with the ground truth setup.

To capture the ground truth data, we built a hyperspectral camera based on a VariSpec tunable color filter. The hyperspectral image is captured time-sequentially. We use a flip mirror to bypass light into this reference system. This way, it sees the object from the same point of view as the CTIS system. The VariSpec filter has a bandwidth of 7 nm. We therefore capture our scenes in 7 nm steps and also reconstruct the CTIS images with this channel width. The camera captures the scene with a spatial resolution that is around $\times 4$ higher (in each dimension) than that of the zeroth order image of the CTIS prototype. Figure 4(b) shows a modulation transfer function (MTF) of the CTIS system compared to the ground truth system. This has been determined using a measurement of a Siemens star. It shows that we have a three times better imaging quality with the ground truth system than with the CTIS system (zeroth order). It thus can be used to train our network accounting for super-resolution.

4 Experimental Results

Synthetic Data Spectral reconstruction, as well as semantic segmentation results, are presented in this section. To highlight the contribution of spectral information for object classification, we compare results obtained by training the ResUnet using the reconstructed hyperspectral cubes from CTIS measurements with the ones obtained using RGB images extracted from the reconstructed hyperspectral cubes. Quantitative results are shown in Tables 1 and 2, while the qualitative are in Figures 5 and 6. From Table 1 and Figure 5 it can be seen that the

Table 1: Quantitative metrics for spectral reconstruction and image super-resolution on FVgNet [10].

Split	PSNR (dB)	SSIM	MAE ($1e^{-3}$)
Train	51.943	0.995	1.5
Test	51.781	0.995	1.6

Table 2: Quantitative metrics for semantic segmentation on the test set of FVgNet [10]: *Obj* refer to the semantic segmentation task on object classes meanwhile *R/F* refer to the task of classifying real and fake objects (better in bold).

Input	mIoU (%)		F1		Precision		Recall	
	Obj	R/F	Obj	R/F	Obj	R/F	Obj	R/F
RGB	78.61	91.54	0.878	0.953	0.862	**0.966**	0.907	0.941
Hyperspectral	**86.63**	**91.95**	**0.926**	**0.956**	**0.902**	0.958	**0.957**	**0.954**

model produces acceptable reconstructions both spatial and spectral-wise with $\times 5$ super-resolution factor. Figure 5 shows how semantic segmentation using only RGB data fails sometimes to learn correct pixel labels due to the limited information carried out by the three color components, instead the network might rely heavily on semantic cues. In the case of semantic segmentation from spectral data, results are much better for both classification tasks, in particular achieving a gain of more than 8% on the objects' semantic segmentation task. Although segmentation metrics for Real/Fake classification task using spectral data is only slightly better than the one using RGB as shown in Table 2 and Figure 6, such behavior can be due to the network capability to better leverage semantic cues in the latter case.

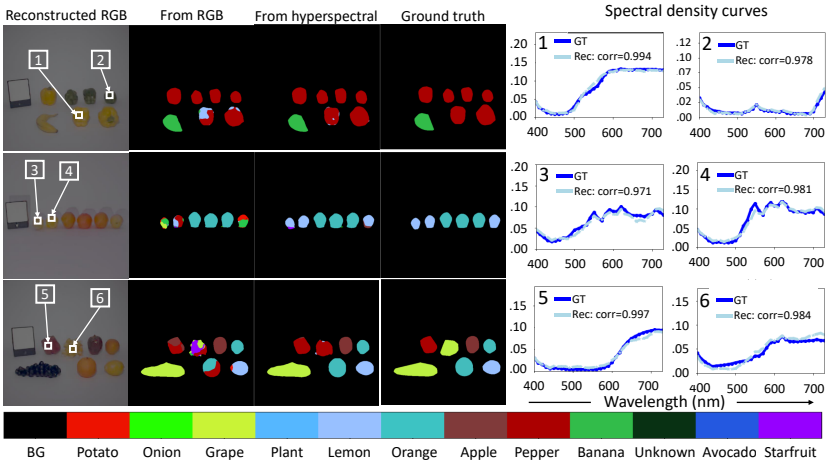


Figure 5: Qualitative results on hyperspectral reconstruction and semantic segmentation of various objects. We show also spectral density curves of some chosen image regions.

Real Data In this section we present reconstruction results on real data captured by our compact CTIS system. Figure 7 shows a few reconstructed images in sRGB space and some selected individual spectral bands along with spectral density curves of some image regions to highlight the discrepancies between the spectrum of real and fake red peppers.

5 Conclusion

We presented a compact CTIS prototype using a Galilean design and a ground truth acquisition apparatus that allows to capture high quality hyperspectral images. We showcased spectral reconstruction and material classification capability from CTIS measurements using a deep learning based approach to reconstruct spatially super-resolved hyperspectral cubes and perform semantic segmentation of fake and real fruits and vegetables leveraging their spectral signature.

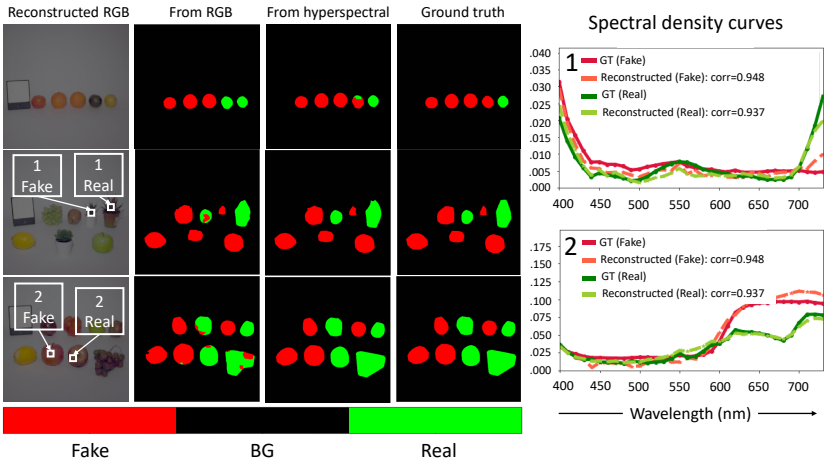


Figure 6: Qualitative results on Real/Fake semantic segmentation. We also show spectral density curves of some chosen image regions.

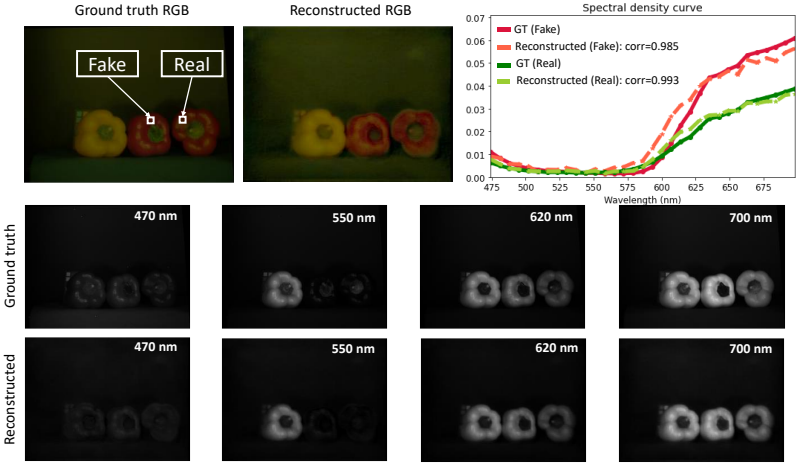


Figure 7: Qualitative reconstruction of a real CTIS scene containing real and fake red peppers. The reconstruction image has $\times 2$ the resolution of the zeroth diffraction order.

References

1. M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14 118–14 129, 2018.
2. T. Okamoto and I. Yamaguchi, "Simultaneous acquisition of spectral image information," *Optics Letters*, vol. 16, no. 16, p. 1277, aug 1991.
3. N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Optical Engineering*, vol. 52, no. 9, p. 090901, sep 2013.
4. Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3048–3059, oct 2015.
5. C. E. Volin, "Portable snapshot infrared imaging spectrometer," Ph.D. dissertation, The University of Arizona, 2000.
6. M. J. Ahlebæk, M. S. Peters, W.-C. Huang, M. T. Frandsen, R. L. Eriksen, and B. Jørgensen, "The hybrid approach–convolutional neural networks and expectation maximization algorithm–for tomographic reconstruction of hyperspectral images," *arXiv preprint arXiv:2205.15772*, 2022.
7. M. Zimmermann, S. Amann, M. Mel, T. Haist, and A. Gatto, "Deep learning-based hyperspectral image reconstruction from emulated and real computed tomography imaging spectrometer data," *Optical Engineering*, vol. 61, no. 5, p. 053103, 2022.
8. M. Mel, A. Gatto, and P. Zanuttigh, "Joint reconstruction and super resolution of hyper-spectral ctis images," in *33rd British Machine Vision Conference 2022*. BMVA Press, 2022.
9. F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020.
10. M. Makarenko, A. Burguete-Lopez, Q. Wang, F. Getman, S. Giancola, B. Ghanem, and A. Fratalocchi, "Real-time hyperspectral imaging in hardware via trained metasurface encoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 12 692–12 702.
11. S. Amann, T. Haist, A. Gatto, M. Kamm, and A. Herkommer, "Design and realization of a miniaturized high resolution computed tomography imaging spectrometer," *EPJ Web of Conferences*, vol. 266, p. 02001, 2022.