HYPERBOT – A BENCHMARKING TESTBED FOR ACQUISITION OF ROBOT-CENTRIC HYPERSPECTRAL SCENE AND IN-HAND OBJECT DATA

Nathaniel Hanson[1*], *Tarik Kelestemur*[1], *Joseph Berman*[1], *Dominik Ritzenhoff*[1], *Taskin Padir*[1]

[1]Institute for Experiential Robotics, Northeastern University, Boston, Massachusetts, USA

ABSTRACT

Robots will benefit from identifying novel objects in their environments through multi-modal sensing capabilities. The overarching goal of this research is to accelerate multi-modal sensor data collection for general-purpose robots to infer material properties of objects they interact with. To this end, we designed a benchmarking testbed to enable a robot manipulator to perceive spectral and spatial characteristics of scene items. Our design includes the use of a push broom Visible to Near Infrared (VNIR) hyperspectral camera, co-aligned with a depth camera. This system enables the robot to process and segment spectral characteristics of items in a larger spatial scene. For more targeted item manipulation, we integrate a VNIR spectrometer into the fingertips of a gripper. By acquiring spectral signatures both at a distance and at grasp time, the robot can quickly correlate data from the two sensors, each of which contain distinct quantum efficiencies and noise. Our approach to this challenge is a step towards using spectral data for enhanced grasp selection in cluttered environments and automated ground-truthing of hyperspectral sensor data. This paper describes our approach to the design of this benchmarking testbed. The project code and material list are located here: `https://github.com/RIVeR-Lab/HyperBot`.

***Index Terms*—** Hyperspectral Imaging, Robot Spectroscopy, Grasp Planning, Sensor Fusion, Ground Truth Acquisition

1. INTRODUCTION

Hyperspectral imaging (HSI) has traditionally been applied in the context of remotely sensed images, where limited spatial resolution fostered the development of multi-band sensors to compensate for the coarseness of ground data. Existing datasets such as Pavia University and Indian Pines are exhaustively studied and are viewed as benchmarks for nearly every new development in HSI [1]. Despite the popularity of hyperspectral imaging, the time and effort required to create accurate ground truth measurements is a commonly cited problem in expanding available datasets [2]. In supervised machine learning techniques, labels for a portion of the data are required to both train and evaluate the performance (accuracy, precision, recall) of models. Labeling hyperspectral datasets has traditionally been undertaken in the laborious process of a technician carrying a field spectrometer and acquiring spectral signatures pinpointed to a referenced coordinate sample. Our work seeks to automate that process, by developing a first of its kind robot workcell capable of collecting hyperspectral datacubes for a scanned set of items and associating camera pixels with high resolution spectral readings from in-hand sensor. Our process allows for the rapid identification of items in a table scene through registered HSI and point cloud measurements. The contributions of this paper are:

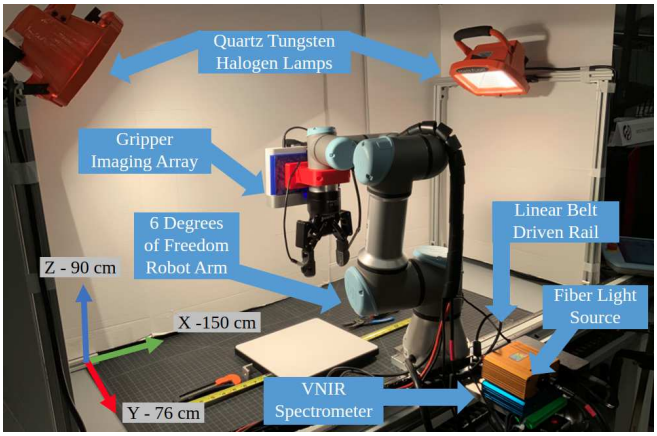

Fig. 1. Spectral benchmarking testbed with architecture to acquire scene hyperspectral datacubes and in-hand spectral data. Note the axes orientations and cell dimensions.

1. The design, analysis, and procedures for an automated hyperspectral ground truthing testbed.

2. A pipeline to process and convert spectral readings between a VNIR hyperspectral camera and fiber optic spectrometer in real time.

Developing this setup enables the automated collection of hyperspectral data, without the need for operator intervention. This work is also needed to enable future research on the role hyperspectral imaging can play in enhancing object handling and perception for autonomous systems.

This research is supported by the National Science Foundation under Award Number 1928654.

*Corresponding author `hanson.n@northeastern.edu`

2. RELATED WORK

Prior work in use of spectral data in robotics can be split into two main approaches. The first involves industrial processing of products on a conveyor belt as demonstrated by [3], [4]. These applications rely on spectral identification of objects using pretrained machine learning models to enable planar motion planning.

The second use case is abstract object recognition in unstructured environments. In our prior work, we focused on the use of spectral signatures in terrain classification for autonomous vehicles [5], and grasped object manipulation via the use of point-based spectrometers [6]. Erickson et al. [7] has demonstrated enhanced material recognition with scanning spectrometers mounted on a mobile manipulator; however, their work heavily focuses on common objects likely to be encountered in everyday life.

Although there is demonstrable value in understanding physio-chemical material characteristics with spectrometers, the close working distance and limited spatial coverage of the photodetector array requires motion planning to bring the gripper into the correct 6-dimensional pose to acquire a spectral measurement which is time and energy intensive for robots. Other projects have attempted to design specialized grasping systems for fruit ripeness detection [8], but such a manipulator does not translate to the diversity of grasps required by general-purpose robots.

3. ROBOT DESIGN

The design of the testbed emphasizes the repeatability and accuracy of the system, especially when acquiring spectral measurements. The testbed employs a 6 degrees of freedom (DoF) robot manipulator (Universal Robotics UR3e). The end-effector is a 2-fingered gripper (RobotIQ 2F-85) which is controllable through a serial interface. The robot system interfaces with a Linux PC running Ubuntu 20 and the open source Robot Operating System (ROS) [9].

The arm is mounted to a linear actuator (LOPRO) which is driven by a stepper motor connected through a serial USB connector. The drive motor can move the rail gantry at a speed from 0.01-1 m/s. This rail in effect provides a seventh degree of freedom, increasing the workspace of the arm and enabling scanning for the hyperspectral camera as further detailed in Section 4. The environment is isolated from the rest of the ambient room lighting by an extruded aluminium frame which provides additional mounting points for sensors and active lighting. We utilize two Quartz Tungsten Halogen (QTH) lighting sources to provide diffuse, full-spectrum illumination within the testbed. The complete setup is shown in Fig. 1.

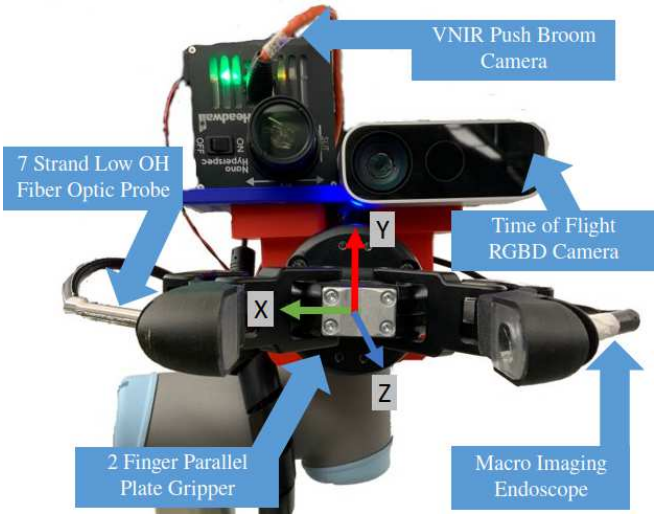

Fig. 2. Multi-modal perception setup for sensing both spatial-spectral characteristics of total workspace and in-hand. *NB*: Z-axis is perpendicular, coming out of the page.

4. SENSING SYSTEM

Our approach to multi-modal sensing combines two classical modalities extensively used in remote sensing: VNIR push broom hyperspectral sensing and point clouds. Most remote sensing point cloud datasets are generated with Light Detection and Ranging (LIDAR) sensors [10]; however, in our design we substitute a Time-Of-Flight (ToF) camera (Azure Kinect) to provide both point clouds and RGB images. As an active sensor, the ToF camera emits infrared light beyond the VNIR range, meaning spectral readings are not perturbed by the operation of this sensor. Depth and RGB images are captured at a consistent 30 frames per second (fps).

The Hyperspectral Camera (Hyperspec-Nano) is mounted with its lens aligned with the ToF camera. A single measurement of the camera consists of 640 pixels each containing 273 wavelength channels between 400 - 1000 nm. The camera is run in 12-bit depth mode for all the experiments in this work. To create the composite datacube, streamed line images are passed into a dynamically allocated buffer as the rail translates along the x-axis. The exposure time is set to 25 ms and the system moves at 0.021 m/s.

The gripper contains additional sensors placed behind protective elastomer gel (Solaris) which provides added tack for grasping in-hand objects. The optical clarity of the gel membrane was confirmed in our previously conducted transmission tests [6]. The reinforced finger tips are 3D printed with reinforced carbon fiber for added rigidity, and contain clamp rings to minimize sensor movement. One finger contains an endoscope with a Light Emitting Diode (LED) light ring to capture macro texture images of grasped items. This sensor is unused in the scope of this paper, but will be included in future work. The opposite finger contains a fiber

optic probe with 8 Low OH fiber optic cores - 7 provide illumination from a QTH light source and the central one reads the signal back to a VNIR spectrometer (StellarNet). Data from the spectrometer is acquired at 10 Hz, and consists of a 2048 length vector of photon counts acquired during the integration period in the wavelength range of 350 - 1150 nm. The complete sensing array is detailed in Fig. 2.

5. SPECTRAL MATCHING

A core capability of our system is the ability to associate spectral signatures sensed by the Hyperspectral camera, with signatures acquired in the gripper. Both sensors utilize different grating technologies and photodetectors with distinct quantum efficiencies, making the transformation more complex.

5.1. Reflectance Calibration

To begin, both devices are calibrated to normalized reflectance values. The in-hand spectrometer was used to grasp a Spectralon reflectance standard in 10 trials in different gripper orientations. The probe was then capped and 10 dark current readings were acquired. Eq. 1 shows the procedure to normalize readings to a range of [0,1]. Readings for the StellarNet spectrometer were very low noise so the mean value of the dark and light calibrations is used here. L is a bright calibration signal, D is the dark signal, and S is a single time step input input signal. Fig. 3 shows a collection of calibrated spectral signatures.

$$\mathrm{S}_{\text{cal spec}} = \frac{\mathrm{S}_{\text{spec}} - \overline{\mathrm{D}}_{\text{spec}}}{\overline{\mathrm{L}}_{\text{spec}} - \overline{\mathrm{D}}_{\text{spec}}} \tag{1}$$

Similarly, a sample datacube was acquired using the previously set parameters for the hyperspectral camera. The same Spectralon captured the reflectance values for 100 frames and 100 capped frames provided the dark current readings. The data was noisier than the spectrometer with occasional dead pixel readings. To prevent these from skewing the results, the max and min of the signals are taken in Eqn. 2.

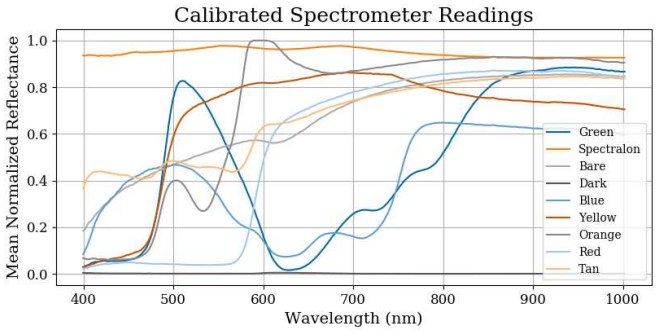

Fig. 3. Reflectance calibrated spectrometer readings using colored blocks for evaluation of VNIR spectral signatures.

$$\mathrm{S}_{\text{cal HSI}} = \frac{\mathrm{S}_{\text{raw}} - \min(\mathrm{D}_{\text{HSI}})}{\max(\mathrm{L}_{\text{HSI}}) - \min(\mathrm{D}_{\text{HSI}})} \tag{2}$$

5.2. Downsampling

The camera and the spectrometer also have slightly different wavelength ranges. To account for the difference here, we clip the input spectrometer signal to the inclusive wavelength range of the spectral camera. The dimensionality of the HSI signal is an order of magnitude lower than that of the spectrometer. To intelligently reduce the dimensionality of the spectrometer signal to match that of the camera, we consider the Full Width Half Maximum (FWHM) values for the photodetector. For the camera, this value is 6 nm. To downsample the spectral values, we pass a moving Gaussian filter over the spectrometer data using the derivation from [11]. Using each wave in the HSI range, λ_{hsi}, centered on the target wavelength from the spectrometer, we weight the signal contributions using:

$$\text{Weight}(\lambda) = \frac{1}{\frac{FWHM}{2.355}\sqrt{2\pi}} \exp\left(\frac{(\lambda_{hsi} - \lambda_{spec})^2}{2\left(\frac{FWHM}{2.355}\right)^2}\right) \tag{3}$$

5.3. Detector Compensation

As a final step, our model must account for differences in optics efficiencies. For instance, the silica in the fiber optic cable introduces signal loss at known wavelengths [12]. For this processing step, real non-zero and non-max signals were required. We collected a sample hyperspectral datacube and spectral readings from grasped painted wooden blocks. These simple items were chosen since they covered the visible light spectrum. The HSI was segmented into rectangular bounding boxes encompassing the spectral signatures of the region of each color type.

We solve a multivariate regression problem to find corrective weight factors at each wavelength of light with Non-negative Least Squares [13]. Namely, we take a matrix B of sampled spectral camera readings. As an input to the system, we stack multiple spectrometer readings that have been reduced according to the steps in Sections 5.1 and 5.2. Solving a linear regression problem yields a mapping $A \rightarrow B$ with 273 factors, forming an appropriate translation between sensors. To avoid overfitting the model, the collected sample values are randomly shuffled into test-train sets with a ratio of 1:3, respectively. The mean squared error by wavelength is plotted in Fig. 4.

6. OBJECT SEGMENTATION

The second part of our work discusses the procedure needed to locate the 3D coordinates and align them with an HSI datacube for proper grasping.

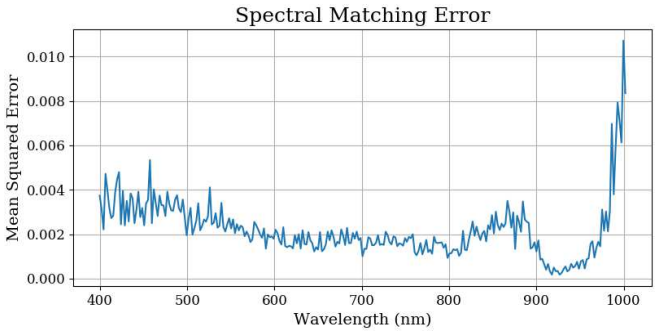

Fig. 4. Mean squared error for fit linear regression model. Note the higher noise at longer wavelengths is correlated with increased sensor noise in the hyperspectral camera.

6.1. Image Registration

After a datacube is collected with a single linear scan of the workspace area, the rail is commanded to bring the arm back into the center of the workspace environment. The camera end effector is rotated to align the lens perpendicular to the table. Here the field of view of the ToF camera, 90° horizontal and 59° vertical, captures the scene with a single RGB image.

With knowledge of the forward and inverse kinematics of the robot arm, we can rotate the view of the camera, T_c to that of the base linkage of the robot arm T_{base} [14]. This step enables all camera readings to be localized in the proper global coordinate frame. The same transform procedure is applied to the point cloud readings.

To register the hyperspectral datacube, three channels representing RGB values are composited. The image is converted to grayscale to aid in the generation of image keypoints and descriptors, which are selected using the SURF algorithm [15]. Because the hyperspectral spatial resolution is much coarser than the high-resolution image from the Kinect, we select a large quantity of features, 2,000, to describe the components in the scene. To handle cases where the visual characteristics of an scene are homogeneous, we add fiducial markers to ensure some scene contrast [16]. Because the images are known to lie on the same plane, we can use a homography to warp the hyperspectral image into the coordinate frame of the ToF RGB image. To prevent noise and diminished resolution from perturbing the solution, RANSAC [17] is used to exclude false matches from the homography calculation.

6.2. Grasp Planning

Our grasp planning framework starts by clustering the objects in the workspace. To do so, we first segment the table plane from the raw point cloud using the RANSAC method. Next, the point cloud above the table is extracted and fed into a Euclidean clustering algorithm which generates the individual point clouds of the objects. Finally, we apply Principal Component Analysis (PCA) to find the 6D poses of the objects $T_i \in \mathrm{SE}(3)$. This point cloud processing pipeline is implemented using the PCL library [18]. Fig. 5 shows the point cloud segmentation and calculated object axes.

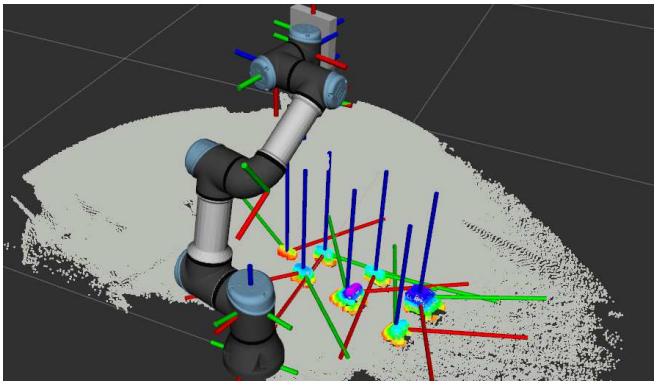

Fig. 5. Real time robot visualization of point cloud data and table scene objects. The centroid of each cluster is associated with a local coordinate frame.

Once we have the poses of the objects, we then use a simple heuristic-based top-down grasp detection strategy. This strategy respects the added mass and size of the sensors mounted near the end effector. The z-axis (downward) of the grasp position is equal to the height of the object plus a pre-determined offset and the x, y-axis is equal to the object's centroid position. For orientation, we set the z-axis to the negative plane normal direction and y-axis to the largest segment of the object. From this, the x-axis is found as the cross product of the y and z axes. To move the robot to the desired grasp pose, the inverse kinematic problem (IK) is solved using the Levenberg–Marquardt algorithm [19]. Finally, the found joint angles from the IK solution is sent to the UR3e's low-level position controller.

Once the object gripper is at its grasp planned position, the fingers are commanded to close; commencing the collection of readings with the spectrometer. From this point, objects can either be released or removed from the scene. The end-to-end cycle of acquiring a datacube takes approximately 1.75 minutes, while the grasp planning and execution takes an average of 7 seconds per object.

7. CONCLUSION

Our work demonstrates the efficacy of a first of its kind robot cell enabling the collection and ground truthing of hyperspectral data. Our methodology allows a multi-degree of freedom robot arm to capture hyperspectral datacubes, point clouds, and in-hand spectral readings.

In future iterations of this work, we plan to orient the camera in different positions with the arm, enabling the acquisition of hyperspectral data from multiple angles relative to scene items. This will enable the generation of dense spectral-spatial models of object points not otherwise visible from a top-down perspective. This extension could logically lead to

3D reconstruction of objects and the interrogation of the spectral properties at each point.

While this work focused on the calibration, alignment, and association of spectral data, our future work lies in understanding how hyperspectral data can enable robots to better manipulate unknown objects. The authors hope the open-sourcing of the control and processing code will encourage the promising integration of spectral sensing into robotics.

8. REFERENCES

[1] Konstantinos Makantasis, Konstantinos Karantzalos, Anastasios Doulamis, and Nikolaos Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 4959–4962.

[2] Utsav B Gewali, Sildomar T Monteiro, and Eli Saber, "Gaussian processes for vegetation parameter estimation from hyperspectral data with limited ground truth," *Remote Sensing*, vol. 11, no. 13, pp. 1614, 2019.

[3] Yuedong Ku, Jianhong Yang, Huaiying Fang, Wen Xiao, and Jiangteng Zhuang, "Deep learning of grasping detection for a robot used in sorting construction and demolition waste," *Journal of Material Cycles and Waste Management*, vol. 23, no. 1, pp. 84–95, 2021.

[4] Pierre Barnabé, Godefroid Dislaire, Sophie Leroy, and Eric Pirard, "Design and calibration of a two-camera (visible to near-infrared and short-wave infrared) hyperspectral acquisition system for the characterization of metallic alloys from the recycling industry," *Journal of Electronic Imaging*, vol. 24, no. 6, pp. 061115, 2015.

[5] Nathaniel Hanson, Michael Shaham, Deniz Erdogmus, and Taskin Padir, "Vast: Visual and spectral terrain classification in unstructured multi-class environments," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, p. To appear.

[6] Nathaniel Hanson, Hillel Hochsztein, Akshay Vaidya, Kristen Dorsey, and Taskin Padir, "In-hand object recognition with innervated fiber optic spectroscopy for soft grippers," in *2022 5th IEEE International Conference on Soft Robotics, RoboSoft 2022*. IEEE, 2022.

[7] Zackory Erickson, Eliot Xing, Bharat Srirangam, Sonia Chernova, and Charles C Kemp, "Multimodal material classification for robots using spectroscopy and high resolution texture imaging," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10452–10459.

[8] Victoria Cortés, Carlos Blanes, José Blasco, Coral Ortiz, Nuria Aleixos, Martin Mellado, Sergio Cubero, and Pau Talens, "Integration of simultaneous tactile sensing and visible and near-infrared reflectance spectroscopy in a robot gripper for mango quality assessment," *Biosystems Engineering*, vol. 162, pp. 112–123, 2017.

[9] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al., "Ros: an open-source robot operating system," in *ICRA workshop on open source software*. Kobe, Japan, 2009, vol. 3, p. 5.

[10] Wai Yeung Yan, Ahmed Shaker, and Nagwa El-Ashmawy, "Urban land cover classification using airborne lidar data: A review," *Remote Sensing of Environment*, vol. 158, pp. 295–310, 2015.

[11] Eric W. Weisstein, "Gaussian function," From MathWorld–A Wolfram Web Resource.

[12] M Horiguchi and H Osanai, "Spectral losses of low-oh-content optical fibres," *Electronics Letters*, vol. 12, no. 12, pp. 310–312, 1976.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[14] Tully Foote, "tf: The transform library," in *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*. IEEE, 2013, pp. 1–6.

[15] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[16] John Wang and Edwin Olson, "Apriltag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4193–4198.

[17] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[18] Radu Bogdan Rusu and Steve Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4.

[19] Kenneth Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.