Application Note

SPORIAN® MICROSYSTEMS, INC

Machine Learning Enabled Hyperspectral Imaging for Location of Pecan Weevil Larvae among Pecan Nutmeat Pieces

Authors

Kevin F. Harsh
Duc Nguyen
Laura Ferlin
Sporian Microsystems Inc.

Introduction

The pecan weevil, *Curculio caryae* (Horn), can devastate harvests from orchards in multiple pecan-producing states in the US. Pecan weevil is considered the most significant insect pest of pecan producers. If not contained, the pest can dramatically affect the pecan industry's economic impact. Growers can lose 75% or more of an orchards harvest during an infestation [1]. According to the United States Department of Agriculture–National Agricultural Statistics Service, the US pecan production in 2019 was over 265 million pounds with a production value of $469 million.

The weevil's economic importance results from crop loss due to egg laying on the developing pecan nuts and the destruction of the edible nut kernel by the larvae feeding inside the shell. Larvae remaining in the pecans at the time they reach a shelling plant have to be removed, which can be an extremely expensive operation. It is estimated that in many shelling/sorting operations, 13-18% of labor and 8-11% of equipment costs are devoted to the issue of removing the larvae. Often such sorting is done manually, and often ineffectively, as sorting larvae from nutmeat is very challenging because larvae are not highly discernible by physical properties, color and visual appearance.

Current methods using alcohol baths, UV lights, manual inspection tables, and other methods are only partially effective and costly [2].

For example, annual labor and material costs to deal with post-harvest pecan weevil removal are in the range of $60-70 thousand and this can be 'back-breaking' for a small to mid-size shelling operation. Further, pecan weevils that remain in final products and are not removed by sorting can leave a stigma with consumers that lasts for years, preventing future pecan purchases.

The mechanization of pecan processing, including cracking and shelling, is a relatively recent advancement when compared to other mechanized food processing systems. Commercial sorters would require automated inspection systems to identify good pecans from larvae before or after shelling to maintain product quality throughout the entire processing, handling and storage of pecans.

Hyperspectral imaging (HSI), where imaging and spectral scanning are combined to provide spatially represented spectral information, is emerging as a non-destructive, real-time detection tool for industrial and environmental sensing and inspection processes. This is due to its ability to simultaneously obtain large amounts of spatial and spectral information on the objects being studied.

This technology has been applied to a broad range of sensing, agriculture, industrial, and process inspection, food safety, mineralogy, astronomy, chemistry, environmental monitoring, biomedical, and surveillance applications [3,4,5,6] and has made possible the rapid spatial assessment of grade, defects, contamination, and constituent characterization of mixed, raw, and processed materials.

In this application note, we'll explore the use of a low cost hyperspectral measurement system for the real-time, accurate identification and location of pecan weevil larvae among pecan nutmeat pieces.

Experiment/Methods

Data presented here was collected using a Sporian Microsystems SpecIQ™ Hyperspectral Sensor System. The measurement system has the capability to provide real-time radiometric measurement of materials by collecting reflected or emitted light through a slit and breaking the light into spectral components for measurement. Whereas a standard camera only records three distinct colors visible to the human eye, hyperspectral imaging records potentially hundreds of bands of color in a range beyond what the human eye can detect. Key to the system's utility for materials identification is the very wide spectral range captured (350 to 2600 nm), ranging from ultra-violet to short wave infra-red, which allows for the collection of spectral information relating to pigmentation, chemical composition, and microstructure of the material imaged. Image information is generated by spatially scanning the measurement and producing a hyperspectral data cube containing complete X, Y, and wavelength data that can analyzed real time using a smart electronics system deploying machine learning algorithms.

Table 1: Key SpecIQ™ specifications

Spec/Feature	Unit
Wavelength Range	350-2600nm
Spectral Resolution	<10nm
Wavelength Reproducibility	<0.1nm
Wavelength Accuracy	+/-0.2nm
Instrument Dimensions	~4"x4"x3"
Communications	Ethernet
Typical spectra measurement times	<1ms

Key system specifications for the system used in these experiments are shown in Table 1, but Sporian builds and provided similar systems in a range of different formats and configurations for different user applications, including handheld unit, lab instruments, and system for deployment on unmanned aerial vehicles. The system used was a lab configuration, designed to operate connected to a computer, receiving camera settings via a user interface and outputting data to the data cube database file.

Pecan weevil larvae were provided by U.S. Department of Agriculture (USDA) – Agricultural Research Service (ARS), Southern Regional Research Center (SRRC), Food Processing and Sensory Quality Research Unit (FPSQ).

Raw pecan pieces were acquired via a commercial vendor and crushed to an approximate size distribution encompassing the larvae.

All measurements were done under a custom, stable, broadband (UV to IR) light source built from an array of emitter types to cover the wavelength range, driven to balance the light across the measurement spectra, and reflected off of a diffuse reflection coated dome to produce spatially even, directionally diffuse light onto the sample. Calibrated Spectralon® reflectance standards were used to monitor light source stability, but were not used as part of subsequent data processing and analysis.

Initial training data for pecans and larvae were generated using an even layer of material in a sampled dish, which were scanned at many spatial locations, then stirred, then measured again. The training datasets for each included approximately 10,000 individual scans to capture the issue of broken nut and larvae variability.

A drawback of utilizing hyperspectral measurements is that the resulting data can contain very high dimensionality per measurement. One material measurement can contain up to thousands of data points with the spectral information from the sample. Depending on the number of measurements and range of materials the dataset can become very large and complex, and analysis of the data by a human analyst is challenging and time-consuming.

Machine learning (ML), a subset of artificial intelligence, is a data-driven approach to analysis, where instead of fitting a physics-based theory to the data output, an algorithm is used to leverage relationships within the data to generate correlation models. ML is therefore suitable for a situation where the system outputs are unknown, hard to understand, or time-consuming to process by a human analyst. ML models can learn from spectral data, identify patterns, and be able to make decisions with minimal human intervention. The models can rapidly process large amounts of high complexity spectral data to be able to classify unknown chemical compounds and/or determine composition.

All spectral data were processed and analyzed by performing a sequence of pre-processing, dimension reduction, and machine learning classification and/or regression. Preprocessing included the stabilization of signal fluctuations based on illumination source variation and the removal of background signal. Then, principal component analysis (PCA)[7] was performed to find the optimal variance in the data as well as transform the data and reduce its dimensionality. Using the optimized data, the data was then split into a training dataset and a smaller test dataset.

A supervised machine learning model (MLPClassifier via Scikit-learn) was implemented to train on the known dataset, analyzing their relationships. For simple identification of material (nutmeat vs. larvae) a classification model was then used that produced a confidence interval (CI) in the classification of the material spectra. Scikit-learn is an open source machine learning library that supports supervised and unsupervised learning. A regression model could be similarly used for quantitative assessments. Scanned spatial information can then be tied to classification confidence interval to produce a map of material type, in this case larvae versus pecan nutmeat.

Individual larvae were then distributed within a field of pecan pieces. The field was scanned spatially, and an image of the field overlaid on the spatial map of materials to show where the camera and machine learning algorithms predicted larvae were located. In principle, this is done so that an automated sorting system could remove the offending larva.

Results

A total of 49,000 measurements were taken as part of developing the training data for the larvae and pecans in which every 100 measurements were averaged together. Figure 1 shows the averaged data from all the available training data for pecan nutmeat and larvae. As shown, the overall curve of pecans and larvae were very similar, notably in the visible range 400-650nm, which helps explain the difficulty the human eye has distinguishing larvae from nutmeat.

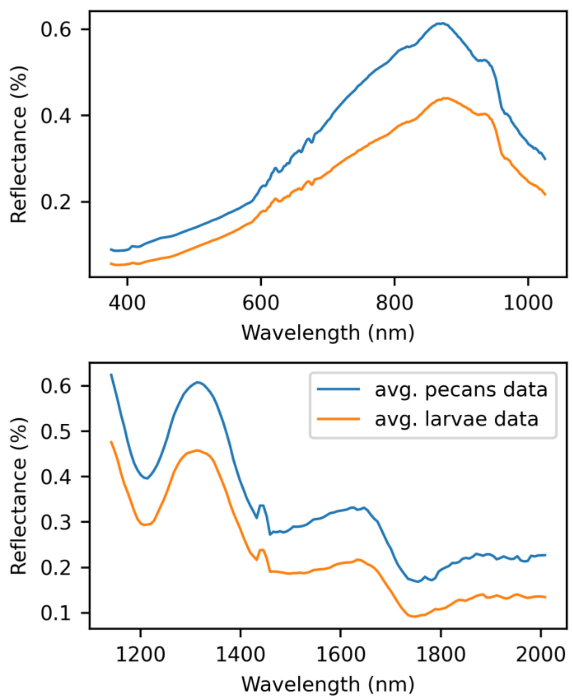

Figure 1: Averaged pecans and larvae reflectance from the training data containing separated pecans and larvae samples

Although there is a shift in reflectance between the two, this simple trend alone was not reliable for classifying between pecan nutmeat and larvae. Even the most stable system measurements are somewhat sensitive to shifting due to subtle lighting changes or system settings changes. Pre-processing of the data, therefore, is beneficial to make the data robust and enable the machine learning to work on the highly varied data in the training and test samples. Pre-processing largely included noise reduction, normalization, and scaling.

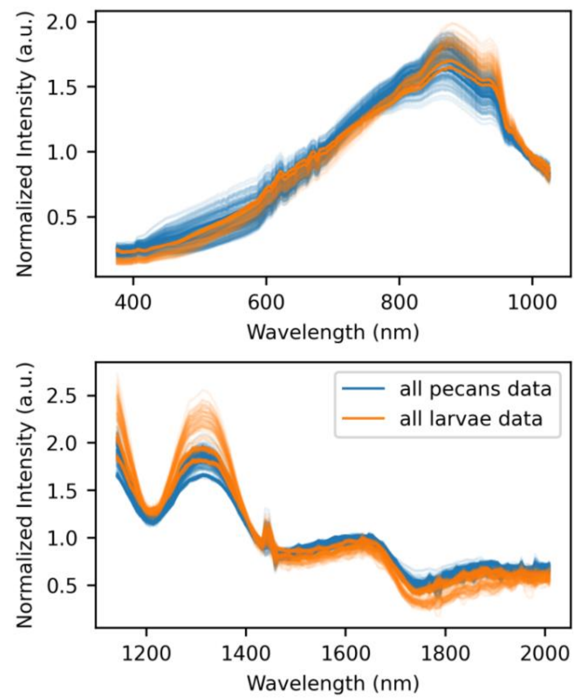

Figure 2: Pre-processed spectral data of pecan nutmeat and larvae after normalization by the mean

Figure 2 shows all of the scans from the training data with pre-processing up to the normalizing step. This normalization was critical in removing variation in the data that are associated with randomness, instability, or use of different settings when scanning complex objects such as pecans and insect larvae. Removing additional variation also allowed the machine learning model to analyze the spectral characteristics, i.e. shape of each measurement, intrinsically. Viewed in this way, the differences between the scans of pecans and larvae become more discernible. After normalization, the data is finally centered and scaled so that the model can best perform its learning on the data. Here, differing trends or even specific pixels in the measurement showing distinctions between the nutmeat and larvae become further evident, and the data is ready for subsequent transform/data reduction.

When using machine learning algorithms, a common step in feature extraction is to separate the data into categories or numerical values of interest. PCA was applied to the training data above to reduce the number of features that the model had to process. This improved the efficiency of the model as it did not have to process as many pixels. PCA is a commonly used technique which transforms the features to new coordinate system(s) optimized to explain as much of the variance as possible among the different features of the data. By finding the features most effective in explaining the data, the method increases interpretability while at the same time minimizing information loss, and the transformation also works as dimensionality reduction. After such processing, the data is reduced to a subset of components that are ranked by those that best explain/convey variance and when plotted the grouping of the data points becomes much more interpretable. Figure 3 shows only three principle components which explained more than 99% of the variance.

A machine learning classification model, such as MLPClassifier, was used to train on the above data, then applied to a 2D scan of a test dish with individual larvae placed throughout. Figure 4-Left is the standard 3 color (CMOS) camera image if the dish of pecan nutmeat with interspersed hard-to-distinguish larvae. Figure 4-Middle, is the same image, with the larvae positions noted. Figure 4-Right, is the result of the 2D scan against training data, with a heat map overlay showing where the classification model predicted a high probability of larvae of 80 to 97.5% for each of the four worms. The classifier model was easily able to highlight larvae locations, which was unsurprising given the separation/distinguishability of the data after feature extraction/transform in Figure 3.

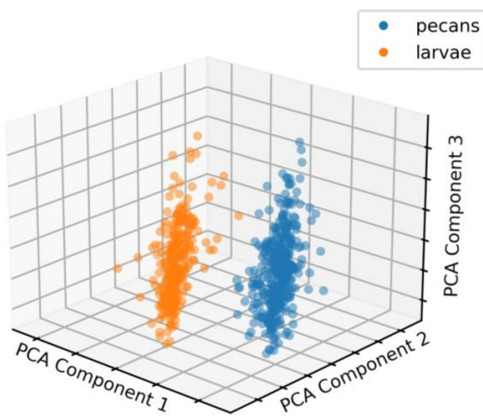

Figure 3: First three features of the data after dimensionality reduction.

Figure 4: (left) Test dish containing pecan nutmeat pieces and four individual larvae; (center) Test dish with positions of larvae manually outlined by rings with dotted lines; (right) Heat-map result of the machine learning model on the test dish, where 'hot' bright yellow is high probability of larvae

Conclusions

Hyperspectral imaging (HSI) is an emerging, non-destructive, real-time detection tool for industrial and environmental sensing and inspection processes, which is being applied to a broad range of industrial applications including agriculture, food processing, and food safety inspection. Sporian Microsystems' hyperspectral measurement technology is an effective tool for the rapid assessment of grade, defects, contamination, and constituent characterization of mixed, raw, and processed materials. The successful use of a low-cost hyperspectral measurement system for the accurate, real-time identification and location of pecan weevil larvae among pecan nutmeat pieces was presented and illustrates how such an instrument could be used more generally within industrial processing environments.

Acknowledgement

Special thanks to Dr. Chris Mattison of USDA-ARS-SRRC-SPSQ, who sourced and provided pecan weevil larvae used here.

1 Hudson, W. G., et al. "Biological control of pecan weevils in the Southeast: a sustainable approach. SARE Fact Sheet Publication# 0 9 AGI 2010." (2010).

2 Weckler, P., et al. "Technologies for enhancing pecan production and processing." International Symposium on Pecans and Other Carya in Indigenous and Managed Systems 1070. 2013.

3 Goetz, Alexander FH. "Imaging spectrometry for remote sensing: Vision to reality in 15 years." Imaging Spectrometry. Vol. 2480. International Society for Optics and Photonics, 1995.

4 Adão, Telmo, et al. "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry." Remote Sensing 9.11 (2017): 1110.

5 Qin, Jianwei, et al. "Line-scan hyperspectral imaging techniques for food safety and quality applications." Applied Sciences 7.2 (2017): 125.

6 Bellon-Maurel, Véronique, and Nathalie Gorretta. "NIR Hyperspectral Imaging for Food and Agricultural Products." Infrared and Raman Spectroscopic Imaging: Second, Completely Revised and Updated Edition (2014): 295-338.

7 Jolliffe, Ian T., and Jorge Cadima. "Principal component analysis: a review and recent developments." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016): 20150202.