5-2019

Counterfeit Detection with Multispectral Imaging

Ian Spatz

Counterfeit Detection with Multispectral Imaging

An Undergraduate Honors Thesis
in the

Department of Electrical Engineering
College of Engineering

by

Ian Spatz

University of Arkansas
May 2019

STATEMENT OF ORIGINALITY

This is to certify that this submission is my own work, and to the best of my knowledge, it contains no materials previously published or written by another person, except where due acknowledgement is made in the paper. Any contribution made to the research by others is explicitly acknowledged in the paper. I also declare that the intellectual content of this paper is the product of my own work, and all assistance received in preparing this paper are acknowledged.

Ian Spatz

ACKNOWLEDGMENTS

This work would not have been possible without the support of the University of Arkansas' Electrical Engineering department and Honors College. I would especially like to thank my research mentor and Electrical Engineering Honors coordinator, Dr. Jingxian Wu, for supporting me through my academic passions and goals. He has always taken time to professionally guide me through my academic interests and undergraduate research career. He consistently provides a helping hand and has exemplified how to be a professional, well rounded engineer as well as a genuine person. I will always be indebted to him. I would also like to give a special thanks to Dr. Adriana Potra for making and supplying the standard earth element solutions used for data collection. A special thanks is also given to my research mentor's Chinese collaborators for supplying multispectral images of agricultural fields, and a special thanks to the Human Environmental Science undergraduate student who made the counterfeit dress. I'd also like to thank my lifelong friend, Reid Sutherland, for being an invaluable computer science resource and helping me install OpenCV on my computer when python2 and python3 library complications arose.

Counterfeit Detection with Multispectral Imaging

By

Ian Spatz

Abstract

Multispectral imaging is becoming more practical for a variety of applications due to its ability to provide hyper specific information through a non-destructive analysis. Multispectral imaging cameras can detect light reflectance from different spectral bands of visible and nonvisible wavelengths. Based on the different amount of band reflectance, information can be deduced on the subject. Counterfeit detection applications of multispectral imaging will be decomposed and analyzed in this thesis. Relations between light reflectance and objects' features will be addressed. The process of the analysis will be broken down to show how this information can be used to provide more insight on the object. This technology provides desired and viable information that can greatly improve multiple fields. For this paper, the multispectral imaging research process of element solution concentrations and counterfeit detection applications of multispectral imaging will be discussed. BaySpec's OCI-M Ultra Compact Multispectral Imager is used for data collection. This camera is capable of capturing light reflectance from wavelengths of 400 – 1000 nm. Further research opportunities of developing self-automated unmanned aerial vehicles for precision agriculture and extending counterfeit detection applications will also be explored.

Table of Contents

List of Figures

1.0 Introduction

With the declining prices of hyperspectral cameras, the applications of these technologies have drastically expanded in multiple industries. The use of hyperspectral cameras is increasingly being employed in the medical [1, 2], pharmaceutical [3], and food industries [4]. Hyperspectral imaging attains both spatial and spectral information through a combination of conventional imagery and spectroscopy. For this image capturing, data is stored in 3-dimensional data sets known as hypercubes, where two dimensions are spatial readings and one dimension is a spectral reading [2]. In the case of hyperspectral imaging, the spectral information is typically from hundreds to thousands of small spectral bands. Multispectral imaging (MSI) is a subcategory of hyperspectral imaging. MSI is also used to capture both spatial and spectral information; however, MSI usually captures spectral information from two to five wavebands that are specific to features of interest. These systems generally provide some combination of visible (400 – 700 nm), near infrared (700 – 1000 nm), short-wave infrared (1000 – 1700 nm), mid-wave infrared (3500 – 5000 nm), and long-wave infrared (8000 – 12000 nm) bands [5]. The use of MSI allows for a quick, non-destructive analysis of objects at reduced costs. MSI cameras are less expensive than HSI cameras, and MSI data requires less processing due to the wider spectral bands.

MSI cameras can be used in a variety of applications. Developing applications include outfitting Unmanned Aerial Vehicles (UAVs) with MSI cameras, providing a cost effective and efficient capability to survey agricultural fields. This paper will explain how MSI data can provide valuable information on agricultural fields. This analysis concept will be applied to other applications including scanning different elemental solutions for classification of element type and concentration and scanning different materials for counterfeit detection. This will be the main focus of this paper. These applications will be decomposed and analyzed to show how different band reflectance measurements will be used to deduce specific and insightful information on the subject. Further research opportunities will also be explored and discussed for these applications. This paper shows how to use BaySpec's OCI-M 8 band MSI camera to analyze elemental solutions, printer documents, and counterfeit clothing.

2.0 Background

2.1 MSI Data Analysis Concept

To understand how MSI data can be indicative of features of interest, Precision agriculture, an information and technology-based farming system, will be examined. Precision agriculture allows for farmers to manage crops, soil, fertilization, irrigation, and other factors, optimally [6]. With an MSI outfitted on a UAV, necessary data for real time assessments on agricultural fields can be used to assist in determining precision agriculture actions to increase crop yield. Currently, farmers typically use average conditions and measurements or years of observant based knowledge to maintain their overall crops. This approach creates variability of crop yield dependent upon the section of the agricultural field [6]. Through the use of MSI equipped UAVs, different sectors of agricultural fields can receive specific amounts of water, fertilizer, herbicides, pesticides, and any other attention required to produce maximum crop yield. However, MSI is only capable of capturing data. Data must be processed so that farmers know how to effectively and efficiently attend to their crop sections. For this research, the main objective will be processing collected MSI data to develop algorithms to allow for real-time data processing. Precision agriculture will be used to exemplify how MSI data can provide hyper specific information for features of interest.

MSI data collected can be used to determine plants' health. There is a correlation between healthy plant structures and light reflectance of different bands. This correlation can be observed through the Normalized Difference Vegetation Index (NDVI). Studies have shown that healthy chlorophyll absorbs the majority of visible light from 400 – 700 nm for the photosynthetic process. Healthy cell structures of leaves also reflect near-infrared light from 700 – 1100 nm. If there are more leaves on a plant, then more wavelengths from 400 – 1100 nm will be affected [7]. Using this knowledge, the standardized NDVI can be calculated using the following equation:

$$NDVI = \frac{(NIR-VIS)}{(NIR+VIS)} \qquad (1)\ [7]$$

where NIR refers to the near-infrared light reflected, and VIS refers to the visible light reflected by the vegetation. The NDVI is a comparison between the difference of visible and near-infrared light reflected to the total of visible and near-infrared light reflected. For this reason, the NDVI can only be between -1 and 1. An NDVI of less than or equal to 0 means that there is no

vegetation while an NDVI of 1 indicates the densest amount of green leaves possible [7]. The use of MSI is apparent when calculating the NDVI of agricultural fields. However, algorithms must be created in order to process the MSI data captured by UAVs.

An example of agricultural field MSI data provided by Chinese collaborators can be seen in Figure 1 below. The MSI captures the normal visible light bands as well as non-visible infrared bands. The non-visible light bands are equated to a visible color. The final image composed is all image bands stacked, forming a single false color image.

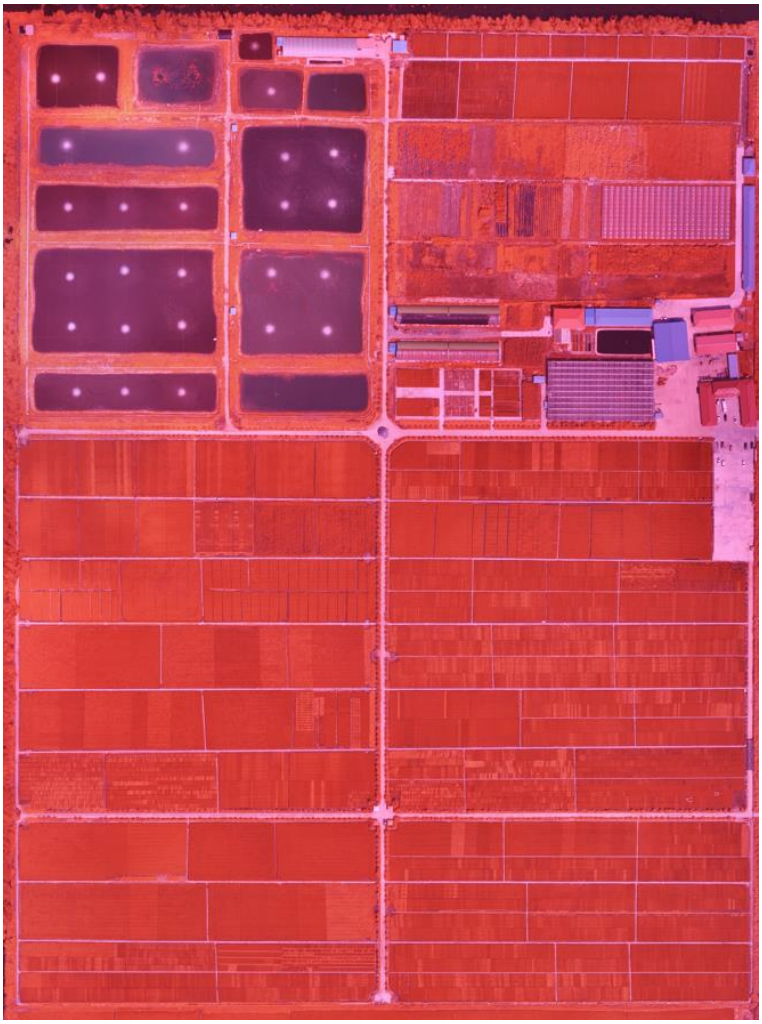

Figure 1. MSI of Agricultural Field (Provided by Chinese collaborators)

Decomposition of the images into separate bands will be required of these images to observe specific types of light frequencies. By comparing these bands, vegetation information can be gathered. The difference of visible light bands and near-infrared bands can then be used to make a calculated estimate on how healthy particular areas of the agricultural fields are. This

data will need to be compared to real-life data taken from samples in the field. By comparing the calculated estimate to the real-life samples, algorithms can be validated so that MSI capturing of agricultural fields can eventually be processed in real-time.

The same MSI analytical concept for precision agriculture can be applied to many different applications. Research has shown that rare earth elements have been used as fertilizer to increase crop yield [8]. For example, mercury solutions have been used to prevent seed-borne illnesses. By scanning different earth element solutions, algorithms can be developed to classify different elemental solutions of varying concentrations in industries' solutions. For this research, BaySpec's 8 band OCI-M Ultra-Compact Multispectral Imager captured MSI data on mercury (Hg), lead (Pb), Neodymium (Nd), and Strontium (Sr) elemental solutions. Table 1 below shows the concentration of each elemental solution used for data collection.

Table 1. Element Solution Concentrations

Element	**Concentration**
Mercury (Hg)	10 parts per million
Lead (Pb)	10 parts per million
Neodymium (Nd)	500 parts per billion
Strontium (Sr)	500 parts per billion

Similar to how the light reflectance in the NDVI correlates to a plant's health, it is likely that these different elemental solutions will have different light reflectance specific to the element. In this paper, the MSI data collected on these solutions will be examined in attempt to classify each solution.

MSI research will also be applied for counterfeit detection. Each printer implements its own digital watermark on every printed page. This digital watermark is known as a Machine Identification Code (MIC). It allows for traceable identification of the specific printer used to print each document [9]. In the case of counterfeit money, not only is it possible to detect counterfeit, but it's also possible to identify the printing source. For this research process, different printers were used to print color images as well as black and white images. The same settings were used to print the documents from different printers. MSI data collected on these printer documents will be analyzed in attempt to classify the printer source of each document.

This research will explore the counterfeit materials and see if MSI data can be used to detect counterfeit clothing. MSI data will be collected on an original manufactured Kate Spade dress and a counterfeit dress. This data will be analyzed to see if there is a significant difference between the light reflectance of each dress.

2.2 MSI Data Capture

The beginning of this research process requires capturing MSI data. Data was captured in a controlled, inside environment. Two halogen lamps were used as the light source. Halogen light sources are essential for MSI data collection due to its continuous spectrum of light output from near ultraviolet (~315 nm) to infrared (~1050 nm).

Data was captured using BaySpec's 8 band OCI-M Ultra-Compact Multispectral Imager. This camera is capable of capturing light reflectance from 400 nm to 1000 nm wavelength using a push broom scanning format. The light reflectance from 400 nm to 1000 nm will be evenly separated into 8 equal bands (band 0 through band 7). This means that band 0 will capture light reflectance from 400 nm – 475nm, and band 7 will capture light reflectance from 925 nm – 1000 nm. BaySpec's CubeCreator software application will then be used to generate hypercubes. It creates bitmap (BMP) image files of each spectral band that correlate the amount of light reflectance to an 8 bit binary number representation between 0 and 255. In order to create a high-quality hypercube, BaySpec's CubeCreator requires a diverse background to correctly stitch the scanned images together. This was achieved by using a newspaper for the background during data collection. Once high-quality hypercubes have been collected, preprocessing steps will be required so that pixel data only comes from specific regions of interest.

2.3 Data Preprocessing

2.3.1 Scale Invariant Feature Transform (SIFT) Keypoint Detection

During the data collection for this research, BaySpec's CubeCreator generated 8 BMP images for each scan. Unfortunately, these images were not perfectly aligned with each other. This was likely caused by human induced error, a common problem during real life data collection processes. In order to overcome this error, image processing took place using Scale Invariant Feature Transform (SIFT) to detect key image features. These features were used to warp the perspective of the image in order to ensure alignment of pixels between band images. This image warping ensures that each pixel between bands can accurately represent 8 spectral readings of the same location.

SIFT works by locating key points of an image through the detection of scale-space extrema. Scale-space extrema identify locations in images that are invariant of scale and orientation [10]. Computing SIFT works by first doubling the width and height of the image. Once this is completed, a Gaussian convolution is used to manipulate the image by blurring the image with a Gaussian function as shown in equation (2) below:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)\ [10]$$

where $L(x, y, \sigma)$ is the output image, $G(x, y, \sigma)$ is a variable scale Gaussian function, '$*$' is the convolution, and $I(x, y)$ is the input image. The Gaussian function, $G(x, y, \sigma)$ is defined by equation (3).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (3)\ [10]$$

This Gaussian convolution is performed multiple times to the image with increasing standard deviation for the convoluted Gaussian function. Once a threshold is reached, the final blurred image will be down sampled to obtain a smaller image. This process is performed multiple times until the picture becomes too small to scale down further. After this process, a scale space of different sizes of the image will be constructed. Next, stable keypoint locations will be detected.

In order to detect stable keypoint locations, the difference-of-Gaussian must be computed. This is done by finding the pixel value difference between the different blurred images. The difference-of-Gaussian, represented by $D(x, y, \sigma)$, can be calculated using equation (4):

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)\ [10]$$

where k is a constant. The $D(x, y, \sigma)$ image shows the pixel value difference between Gaussian blurs. Figure 2 graphically represents this calculation.

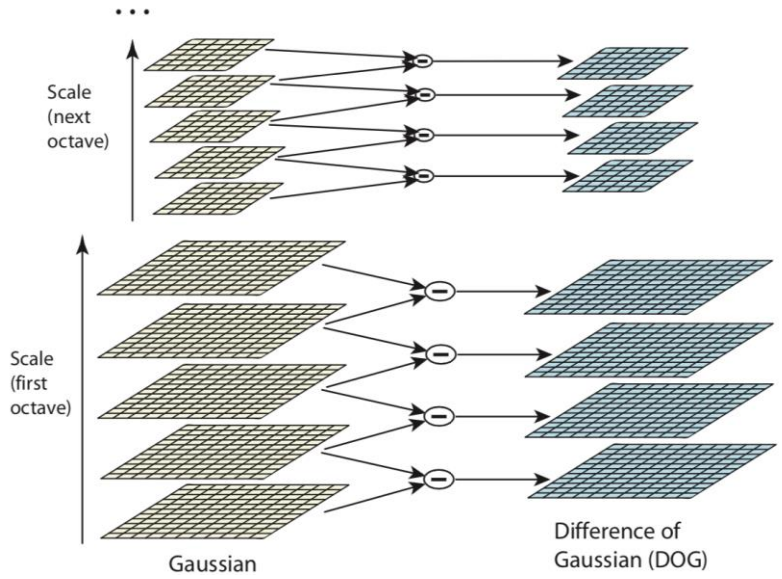

Figure 2. Difference-of-Gaussian Graphic Representation [10]

As you go through the scales, the images become more blurred. This results in a smoothness of pixel value surface intensities. If the surface intensities of the image are smoothed out, then the amplitude of image derivatives decreases as you increase the scale (i.e. increasing σ). This creates a negative proportional relationship between σ and the image derivative. This results in the maximum amplitude image derivative at the beginning of the scale and the minimum amplitude image derivative at the end of the scale. Thus, scale normalization is required to correctly observe the image extrema across the scale.

Theoretically, the scale-normalized Laplacian of Gaussian ($\sigma^2\nabla^2 G$) should be computed for each scale space. Through the normalization of the Laplacian, true scale invariance can be achieved. The extrema of the Laplacian represent the candidates for the keypoint locations. The difference-of-Gaussian is a discrete approximation of the scale normalized Laplacian of Gaussian related by equations (4) and (5).

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \qquad (4)\ [10]$$

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,k)}{k\sigma - \sigma} \qquad (5)\ [10]$$

From these equations, the following can be seen:

$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2 \nabla^2 G \qquad (6)\ [10]$$

From equation (6), it can be seen that the difference-of-Gaussian includes σ^2 required for scale normalization of the Laplacian. The extrema of the difference-of-Gaussian can then be used to detect stable features across the images. The overall SIFT process is represented by Figure 3.

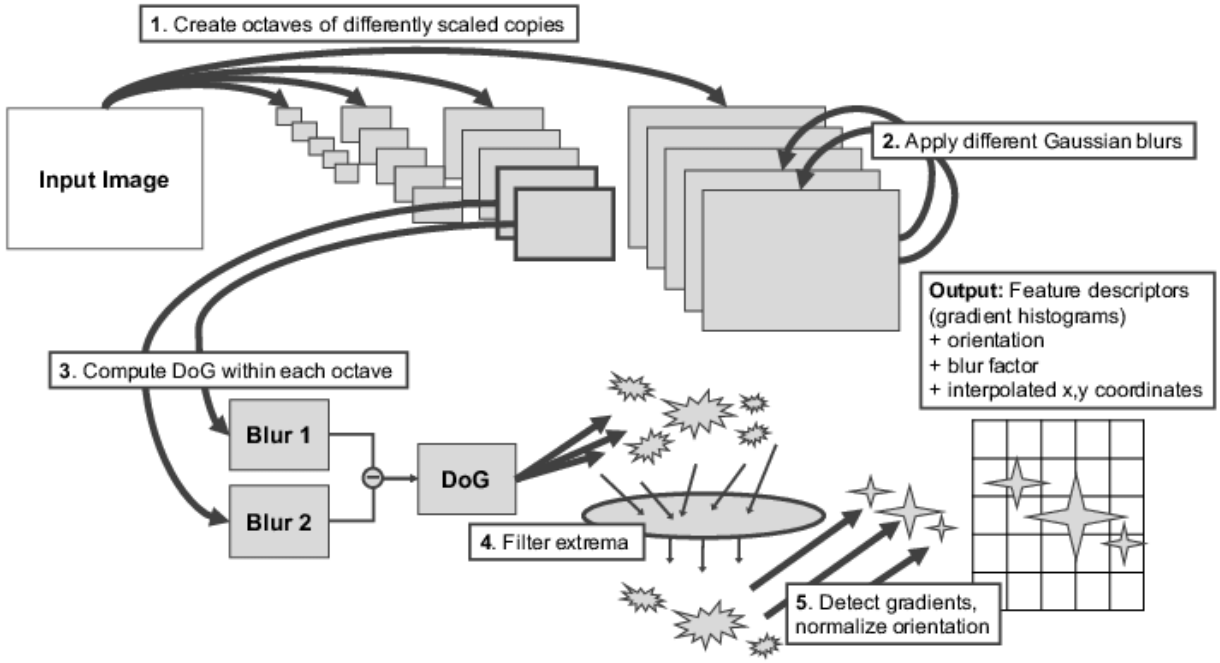

Figure 3. SIFT Algorithm Graphic Representation [11]

Once SIFT features are detected across the band images, the keypoints will be used to warp the perspective image to a base image. By warping all band images to a base perspective, all pixel values should be lined up. This means each spatial location of the images can be analyzed by an 8-dimensional vector of spectral information. SIFT features will be computed with python's OpenCV library.

2.3.2 Image Segmentation

After aligning the band images, image segmentation must be performed on the collected data. This will be completed by applying a binary mask to the hypercubes, allowing for analysis of only the pixels of interest. Pixels of interest for this research include the elemental solutions, the image from the printer documents, and the original and counterfeit clothing. Image segmentation will be performed using MATLAB's image segmenter application.

2.4 Support Vector Machine (SVM) Algorithm Development

Once the data has been preprocessed, algorithm development will take place to see if the MSI data can be used to classify the elemental solutions, printer sources of documents, and

counterfeit clothing. These algorithms will use Support Vector Machine (SVM), a supervised learning model, to classify data with training and testing data. The basic principle of SVM uses support vectors to create planes between different data in order to categorize them. These division planes are known as hyperplanes. SVM uses training data to compute these hyperplanes.

Suppose training data denoted by $\{(x_1,y_1), \ldots ,(x_n,y_n)\}$ is a subset of χ x $\Re$ where χ denotes the space of the input. Therefore, χ exists in a real dimension $\Re^d$. The goal of SVM is to use the training data to compute a function, f(x), that has a maximum deviation of ε from the targets y_i while remaining as flat as possible [12]. A flat function signifies that the derivatives go to 0 at some point. This allows for SVM to classify information as long as errors are within ε. An SVM linear classification function will be broken down to understand the simplest case of SVM classification. The linear function, f(x), is represented by equation (7) below:

$$f(x) = < w, x > + b \qquad (7)\ [12]$$

where < . , .> denotes the dot product in χ. To make this function flat, the norm should be minimized so that $\|w\|^2 = \langle w,w \rangle$. This can be turned into a convex optimization problem as shown in equation (8):

$$\begin{aligned} &minimize && \tfrac{1}{2} \| w \|^2 \\ &subject\ to: && y_i - < w, x_i > - b \ \leq \ \varepsilon \\ &subject\ to: && < w, x_i > + b - y_i \leq \ \varepsilon \end{aligned} \qquad (8)\ [12]$$

assuming that it is possible to approximate all pairs of (x_i, y_i) in f(x) with ε precision [12]. However, it is likely that ε precision cannot be obtained for all pairs. In this case, error can be accounted for by including the slack variables ξ_i and ξ_i^*. The convex optimization problem from before becomes:

$$\begin{aligned} &minimize && \tfrac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\ &subject\ to: && y_i - < w, x_i > - b \ \leq \ \varepsilon + \xi_i \end{aligned} \qquad (9)\ [12]$$

$$subject\ to: \quad < w, x_i > + b - y_i \leq \varepsilon + \xi_i^*$$

$$subject\ to: \quad \xi_i,\ \xi_i^* \geq 0$$

where C is a constant > 0 that is related to the tradeoff between the flatness of f(x) and the toleration of error larger than ε. This type of optimization problem corresponds to an ε-insensitive loss function shown in equation (10) below.

$$|\xi|_\varepsilon := \{0, \quad if\ |\xi| \leq \varepsilon \qquad (10)\ [12]$$

$$|\xi|_\varepsilon := \{|\xi| - \varepsilon, \quad otherwise$$

Figure 4 below shows the linear SVM optimization process.

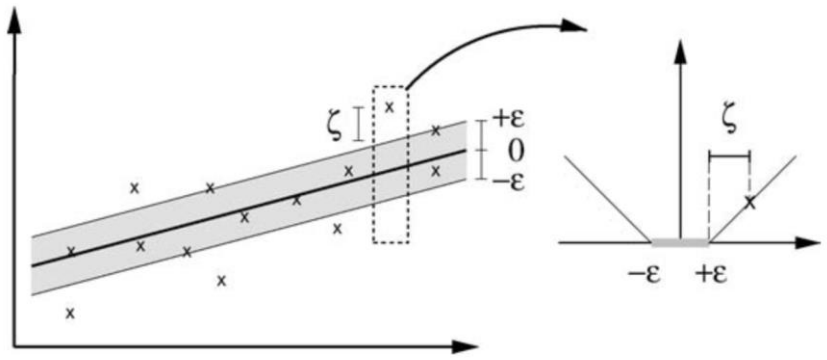

Figure 4. Soft Margin Loss of Linear SVM [12]

The points outside the ε margin are the reason for a non-flat region. Figure 4 displays the basic principle behind SVM. However, the optimization problem from equation (9) can be solved in a dual formulation for nonlinear functions. This means that the previous function will be used as a bound for another function. The previous function will be referred to as the primal objective function, and the Lagrangian will be used for this dual formulation.

A dual set of variables, known as the Lagrangian multipliers, will be required to construct the Lagrangian from the primal function. The Lagrangian, denoted by L, is represented in equation (11):

$$L \coloneqq \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*)$$

$$- \sum_{i=1}^{l} \alpha_i (\varepsilon + \xi_i - y_i + < w, x_i > + b) - \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \xi_i + y_i - < w, x_i > - b) \quad (11)\ [12]$$

where $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are positive Lagrange multipliers used to satisfy the saddle point condition. The saddle point condition means there will be a point on the surface where the partial derivatives are zero with respect to the primal function variables (w, b, ξ_i, ξ_i^*) in the orthogonal direction. In other words, the following conditions will be satisfied.

$$\partial_b L = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 \quad (12)$$

$$\partial_w L = w - \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) x_i = 0 \quad (13)$$

$$\partial_{\xi_i} L = C - \alpha_i - \eta_i = 0 \quad (14)$$

$$\partial_{\xi_i^*} L = C - \alpha_i^* - \eta_i^* = 0 \quad (15)\ [12]$$

The dual optimization problem can be formed by substituting equations (12), (13), (14), and (15) into equation (11). This is shown in equation (16) [12].

$$maximize\ \frac{-1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) < x_i, x_j > - \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i + \alpha_i^*) \quad (16)$$

$$subject\ to \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0\ and\ \alpha_i, \alpha_i^* \in [0, C]$$

From equation (13), the following can be seen:

$$w = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) x_i$$

which results in:

$$f(x) = (\alpha_i^* - \alpha_i) < x_i, x > + b. \quad (17)\ [12]$$

b is a constant that is computed using the Karush-Kuhn-Tucker (KKT) conditions. Since the primal function is convex, the KKT conditions are required for strong duality. More details on how to compute b can be found in *Smola (2004)* [12].

For this research, SVM will be used to categorize the different elemental solutions, printer sources, and counterfeit clothing material. An attempt using Python's scikit-learn or sklearn library was made to perform SVM calculations and categorizations.

3.0 Experiment

3.1 Data Collection

After an understanding of the MSI classification process is obtained, data collection can begin. Data collection was performed on a 10 ppm Pb solution, a 500 ppb Nd solution, a 10 ppm Hg solution, and a 500 ppb Sr solution. These earth element solutions were created by Dr. Adriana Potra from University of Arkansas' Department of Geosciences. Solution samples were transferred to a petri dish where they were scanned in a controlled and inside environment.

Printer documents from different printer sources were also scanned. The image shown in Figure 5 was printed on different printers.

Figure 5. Printed Image [13]

This image was printed in black and white with 600 dots per inch (dpi) on the following printers: HP LaserJet P2055dn, HP OfficeJet Pro 276dw, and Xerox Color C60. The image was also printed in color with 1200 dpi for the following printers: Dell C7765dn, and Xerox Color C60.

Two dresses were scanned for MSI classification. These included an original manufactured Kate Spade dress, and a counterfeit dress created by a student from the University of Arkansas' School of Human Environmental Sciences. All scanned samples were obtained in an inside, controlled environment. Figure 6 below demonstrates the setup of the dress data collection.

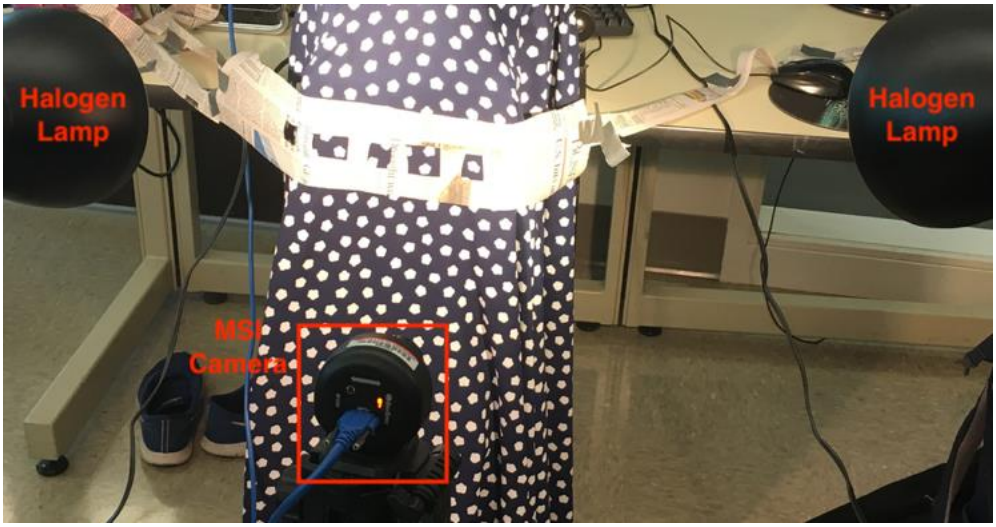

Figure 6. Dress Data Collection Setup

Similar data collection setups were used for each scanned item. The newspaper was used as a diverse background so the CubeCreator software could correctly stich images together.

3.2 Hypercube Generation

After all samples had been scanned, BaySpec's CubeCreator software application was used to generate hypercubes. Each hypercube contains the 8 bands of spectral information. An example of band 4 of the Hg 10 ppm solution is shown in the figure 7 below.

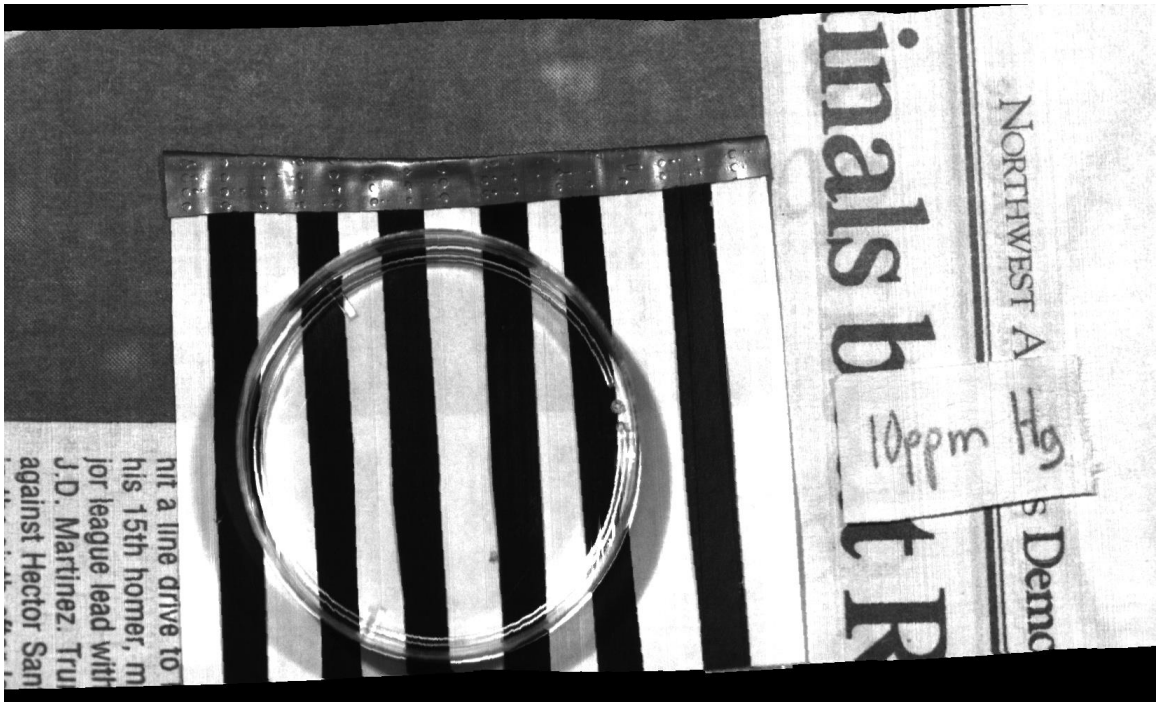

Figure 7. Band 4 of 10 ppm of Hg Elemental Solution

This band represents light reflectance from 700 nm – 775 nm.

3.3 SIFT Keypoint Detection

Once hypercubes were obtained, SIFT feature matching was performed in order to correctly line up the images. Figure 8 exemplifies SIFT keypoint detection using OpenCV's library.

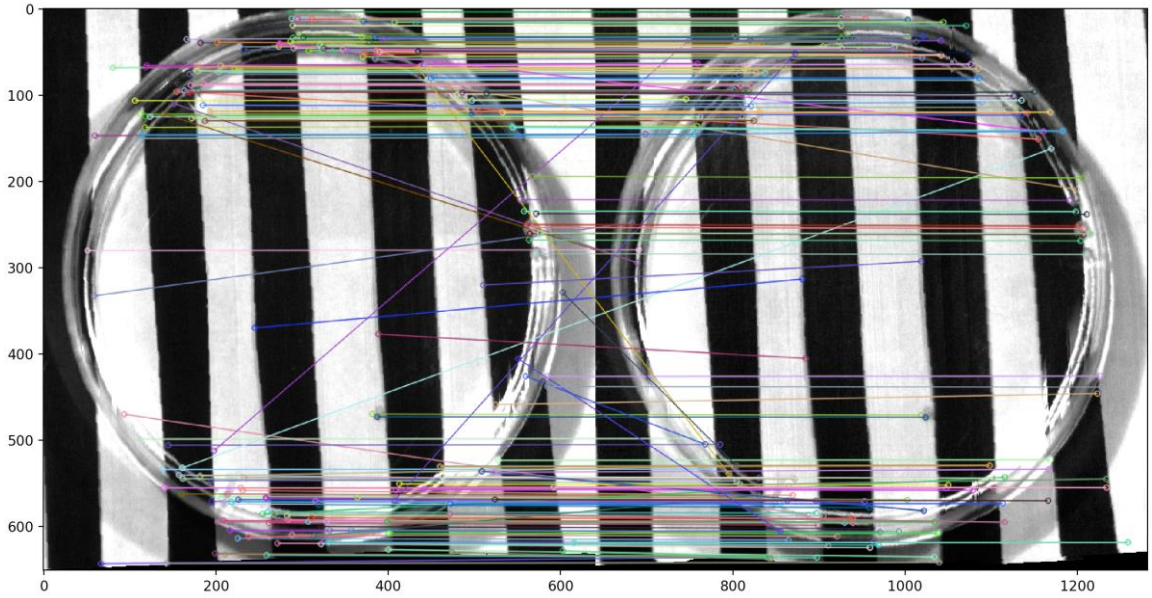

Figure 8. SIFT Keypoint Detection between 10 ppm Hg Band 4 and Band 5

It's apparent that not all keypoints are correct. Each keypoint was manually selected as correct, and only matches on the petri dish's edge were selected. These keypoints were then used to warp the perspective of band 5 to match band 4. This process was repeated for each band using band 4 or band 3 (i.e. a center band) as the base image to map to. By mapping all images to the same base image, all images should align. This ensures that each pixel location will be homologous across each band.

3.4 Image Segmentation

After the images are aligned, segmentation of each image must take place so that only the pixels of interest are analyzed for algorithm development.

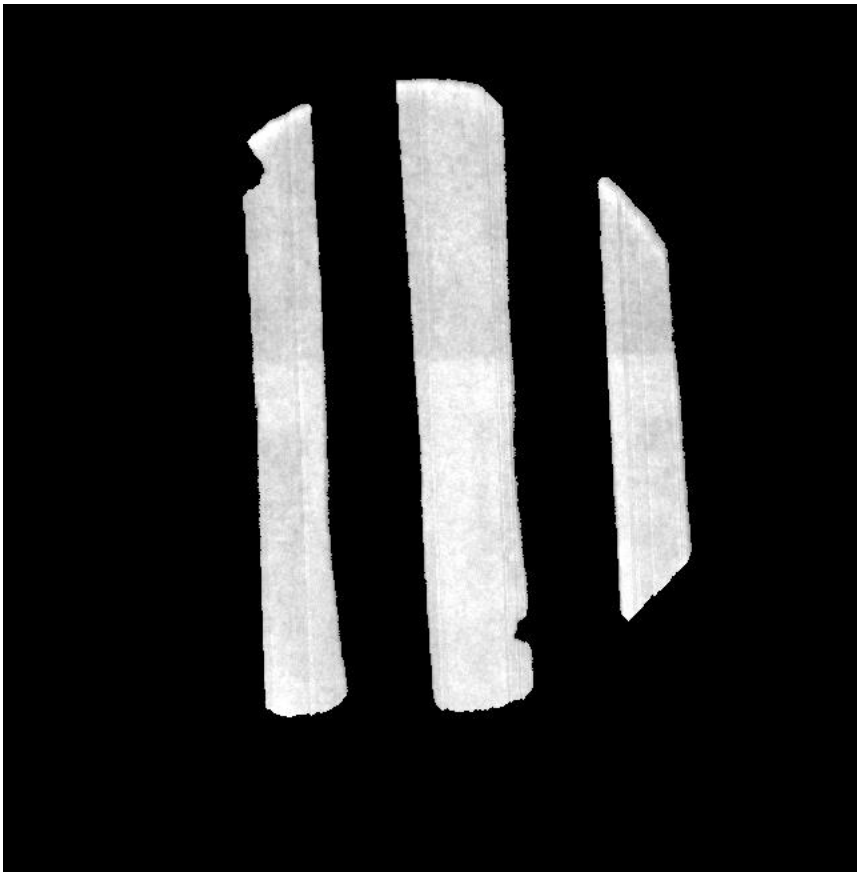

Figure 9. Image Segmentation of 10 ppm Hg Band 4

Once this is complete, these pixels will be combined into a 3-dimensional array. 2 dimensions will represent the spatial location. The 3rd dimension will contain 8 channels, one for each band value. This 3-dimensional array will be the input data for SVM classification.

4.0 Results

This research concluded in segmented images prepared for SVM classification. Image alignment was performed to ensure that spatial positions of pixels were homologous between bands. This allows for a pixel location to be correlated to an 8-dimensional vector represented by each band.

An SVM classification attempt was made using Python's scikit-learn library. However, in order for the SVM classifier to be created, data must be input such that each band image is represented as a single row in a matrix. This will require reshaping the image matrix. The 8 bands of data for each scanned object would correlate to an 8 row matrix where band 0's data

would be in the first row. Labels must also be given for the input data so that the classifier knows how to correlate the training data.

Although the SVM classifier was unable to be obtained, it is likely that MSI data could recognize the printer source of documents if the same image was printed with the same settings. Figure 11 below shows the light reflectance of band 6 from both the HP LaserJet P2055dn and HP OfficeJet Pro 276dw printers.

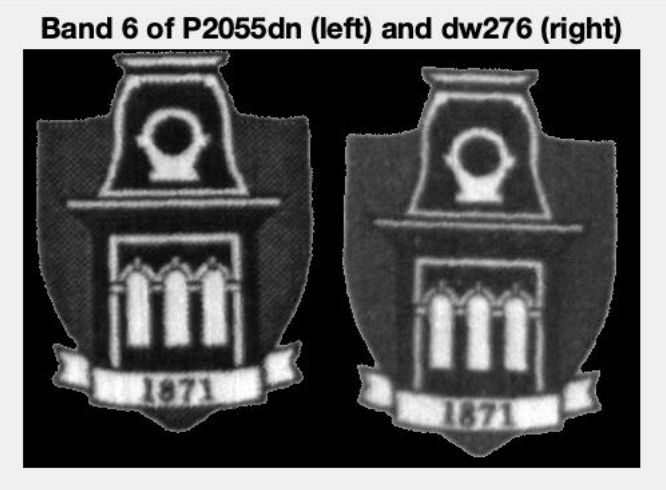

Figure 10. Band 6 Light Reflectance from P2055dn (Left) and 276dw (Right)

If a histogram plot of the pixel values is created, the difference in light reflectance can be seen.

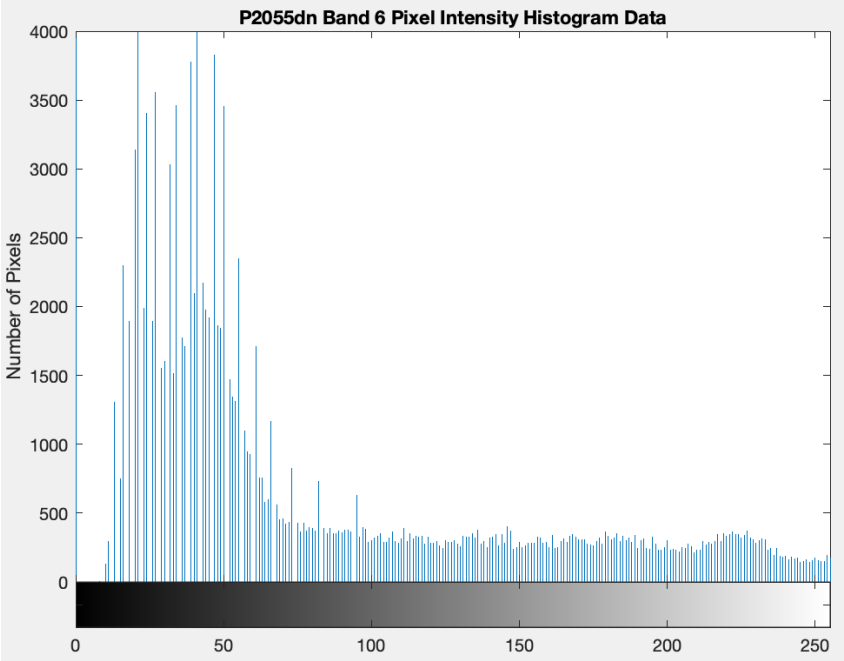

Figure 11. Histogram of P2055dn Band 6 Pixel Intensities

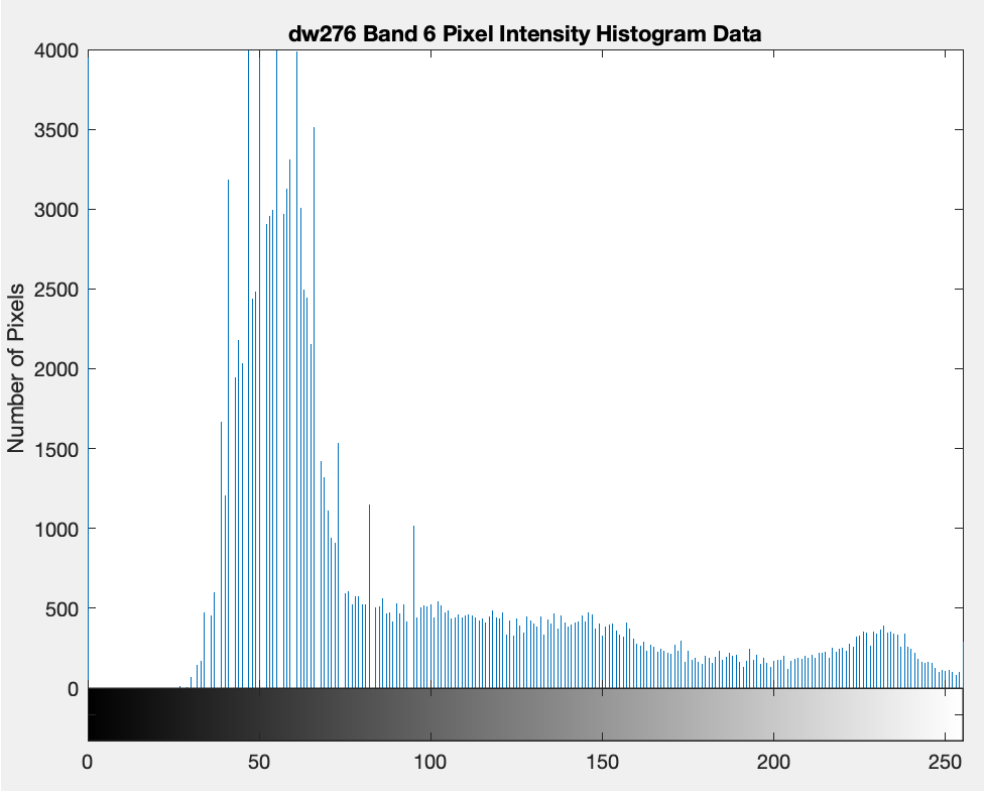

Figure 12. Histogram of dw276 Band 6 Pixel Intensities

The difference in pixel intensities can be viewed through these figures. However, the histogram does not display the pixel location that correlates with these light reflectance intensities. The

SVM classification algorithm would be able to correlate spatial location to light reflectance intensity. A label will be given for each pixel data stating the printer source for each document to train the algorithm.

4.1 Reflection

Originally, it was assumed that the CubeCreator software application would be able to generate hypercubes where every pixel was aligned. However, this did not hold true, and image processing was required to ensure that each pixel location aligned correctly. It is a possibility that human induced error caused the need for image processing. During the data collection process, the camera was set on a tripod while items were moved across the MSI camera's view. The movement of the scanned object did not occur in a perfectly horizontal manner. It would be ideal to have a machine to ensure the object moves only in the horizontal axis. This could be a platform that moves a scanned object, or a machine that moves the MSI camera. It would be ideal if the MSI camera could attach to a machine that could move in the same manner as a 3D printer.

Another adjustment that could be made for better data creation is to have designated markers in the scan to allow for faster alignment during the image processing stage. Four corners could be marked around the scanned object. These corners could be used to easily warp the perspective ensuring pixel alignment across bands.

SVM classification can also be performed in MATLAB. At first, an SVM classification attempt was made using MATLAB. However, python's library was eventually used due to the availability of more SVM example documentation. After obtaining a better understanding of SVM, MATLAB could be a better program to provide SVM classification. The advantage of MATLAB over python is that processed data structures are easily recognizable compared to python. With this advantage, it is likely the data could be structured correctly to create an SVM classification.

5.0 Future Work

To continue with this research, the preprocessed images can be used to create an SVM classification. The data is prepared in a way that could result in a classification if the MSI data is distinguishable. However, one more preprocessing step is required before SVM classification can be made. The final preprocessing step is to reconstruct the image matrix into an 8-row

matrix. Each row will be the input data from each band image. For the elemental solutions, it may be difficult for the SVM classification to recognize the difference when the solution concentration was 10 ppm or 500 ppb. It is possible the SVM classification could, however, pick up different printing sources of documents due to the MIC and the printing technique used by each printer. Different printers usually have different ink cartridges for each printer. These different ink cartridges will likely play a role in the light reflectance measurements. It is also probable the SVM classification could recognize the counterfeit dress due to the counterfeit material used.

5.1 MSI Research Possibilities

Further research possibilities for precision agriculture and counterfeit detection are available for MSI research. Research on MSI data could provide valuable information in agricultural settings. In precision agriculture, further research on the different light reflectance of plants with herbicides and pesticides and ones without could be explored. This would provide more information to farmers and allow them to apply specific amounts of herbicides and pesticides to different sections of the agricultural field. The use of machine learning could also be explored to create self-automated UAVs. Machine learning, also known as artificial intelligence, combines the calculating power of Computer Science with data-based analysis from statistics [14]. Machine learning teaches a computer how to program itself and improve through large amounts of trial and error data. In particular, neural networks will source the self-teaching. Neural networks form complex, non-linear hypotheses from data-based parameters using the computational unit known as a "neuron" [15]. Neurons form an output dependent on multiple inputs. By linking multiple neurons, a multi-layer neural network can be created. Each individual output of a neuron will be an input for all neurons found in the following neural layer. By doing so, information can be used to receive an input image and decompose the picture into separate classes [15]. This will be applied to UAVs for self-automated precision agriculture operation.

After algorithms are developed, applied machine learning could create UAVs that independently survey the unhealthiest field sections first. The real-time information calculated would be processed by machine learning UAVs. The idea is that UAVs could fly at different heights of the field and determine which sections of the field should be analyzed further. This

would prioritize and optimize the path of the UAVs creating the most value out of the UAV operating time.

As for counterfeit detection, research is not limited to just the detection of counterfeit clothing. The same process could be applied to many different items. MSI's nondestructive analysis capabilities would be a highly sought advantage for determining counterfeit.

6.0 Conclusion

Research in multispectral imaging capabilities can allow for a specific, and nondestructive analysis of objects. This can allow for identification of problems that would normally go unnoticed unless drastic measures or intrusions were completed on the objects. Through multispectral imaging, image processing, and algorithm creation techniques, highly desirable information can be obtained. Multispectral imaging allows hyper specific information obtainable to users. For example, the Normalized Difference Vegetation Index is one way to determine the health state of the plants being surveyed, and MSI could be used to further explore this correlation. Multispectral imaging can also allow for counterfeit detection as well as the identification of the original source used for counterfeit.

The detail process for why multispectral imaging is used and how it is used for element solution concentrations and counterfeit detection is explored in this paper. An understanding of how light reflectance of different frequencies is correlated to a specific topic is essential before data can be processed. Analysis of only the specific regions of interest must be completed in order to create a reliable and highly effective algorithm for determining hyper specific information on a topic. Support Vector Machine can be used to determine whether MSI camera data can detect light reflectance differences across a variety of items. This paper shows the preprocessing steps required of MSI data and explains the complete SVM classification process.

References

[1] A. Siddiqi, H. Li, F. Faruque, W. Williams, K. Lai, M. Hughson, S. Bigler, J. Beach and W. Johnson, "Use of hyperspectral imaging to distinguish normal, precancerous, and cancerous cells", *PubMed.gov*, 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/18213691. [Accessed: 19- Nov- 2018].

[2] G. Lu and B. Fei, "Medical hyperspectral imaging: a review", *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, 2014.

[3] A. Palou, J. Cruz, M. Blanco, J. Tomàs, J. Ríos and M. Alcalà, "Determination of drug, excipients and coating distribution in pharmaceutical tablets using NIR-CI", *PubMed.gov*, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5760824/. [Accessed: 19- Nov- 2018].

[4] D. Barbin, G. Elmasry, D. Sun and P. Allen, "Near-infrared hyperspectral imaging for grading and classification of pork", *PubMed.gov*, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/21821367. [Accessed: 19- Nov- 2018].

[5] V. Coffey, "Multispectral Imaging Moves into the Mainstream", *OPN Optics & Photonics News Magazine*, no. 4, pp. 18-24, 2012.

[6] "Precision Agriculture: Tomorrow's Technology for Today's Farmer", *Journal of Food Processing & Technology*, vol. 06, no. 08, 2015.

[7] "Measuring Vegetation (NDVI & EVI)", *Earthobservatory.nasa.gov*, 2000. [Online]. Available: https://earthobservatory.nasa.gov/features/MeasuringVegetation. [Accessed: 12- Oct- 2017].

[8] K. Redling, "Rare Earth Elements in Agriculture with Emphasis on Animal Husbandry", Ph. D, LMU Munich, 2018.

[9] J. Beusekom, F. Shafait and T. Breuel, "Automatic authentication of color laser print-outs using machine identification codes", *ResearchGate*, 2013. [Online]. Available: https://www.researchgate.net/publication/257471976_Automatic_authentication_of_color_laser_print-outs_using_machine_identification_codes. [Accessed: 26- Nov- 2018].

[10] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 1–28, Jan. 2004. [Accessed 20 Feb. 2019.]

[11] Plauth, Max, et al. “Parallel Implementation Strategies for Hierarchical Non-Uniform Memory Access Systems by Example of the Scale-Invariant Feature Transform Algorithm.” *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, May 2016, doi:10.1109/ipdpsw.2016.47. [Accessed 10 Apr. 2019.]

[12] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004. [Accessed 18 Apr. 2019.]

[13] “College of Engineering Electrical Engineering,” *Electrical Engineering | Electrical Engineering | University of Arkansas*. [Online]. Available: https://electrical-engineering.uark.edu/. [Accessed: Mar-2018].

[14] T. Mitchell, "The Discipline of Machine Learning", 2006. [Online]. Available: http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf. [Accessed: 12- Oct- 2017].

[15] "Unsupervised Feature Learning and Deep Learning Tutorial", *Deeplearning.stanford.edu*. [Online]. Available: http://deeplearning.stanford.edu/tutorial/. [Accessed: 12- Oct- 2017].

Appendix:

This shows one example of the data preprocessing done for this research.