Article

A Texture and Polarization Based Imaging Technique for Determining the Relative Water Content of Vegetation

Nicholas Ericksen [1], Romeo Pascone [2]

[1] Nicholas Ericksen; nericksen.student@manhattan.edu
[2] Romeo Pascone; romeo.pascone@manhattan.edu
* Correspondence: nericksen.student@manhattan.edu;

Version May 22, 2019 submitted to Sensors

Abstract: The use of low cost imaging devices for the purpose of agricultural monitoring provides advantages to large scale indoor agricultural operations. These devices permit the use of imaging techniques for the acquisition of physiological indicators useful for the determination of overall plant health. As resources become scarce, the need to minimize the use of agricultural inputs, such as water, will increase. This study presents the design of an efficient-low-cost solution for determining the relative water content of vegetation, experimental implementation, and analysis. The experimental results show the potential capability of using low cost hardware with open source software for the purpose of determining the relative water content of individual plant leaves. The results of performing a linear regression on extracted polarization and texture based features shows a correlation with the relative water content of an Epipremnum Aureum plant.

Keywords: Polarization; Stokes Vector; Texture; Grey Level Co-Occurrence Matrix; Relative Water Content; Linear Regression; Vegetation; Epipremnum Aureum

---

0. Introduction

Implementation of large scale indoor precision agriculture systems will become more prevalent as resources considered as inputs to these systems become increasingly scarce. Monitoring the state of these systems will be crucial for minimizing inputs to the system while maximizing its outputs.

The reduced cost of electronic imaging devices has lowered the barrier of entry into the field of exploratory image analysis and data collection for precision agricultural applications. As large scale deployments of sensors are often costly or lacking spatial temporal resolution, investigations into the use of micro-aerial vehicles, or MAVs, is being investigated by [1] for a reduction in footprint of the physical hardware deployment. Recently Panda et al [2] built upon the work of Tian [3] to increase the reliability of wireless sensor networks at low cost for indoor precision agriculture by including redundant data transfer paths from the sensors to their respective gateways. The importance of open source software and hardware for these types of sensor networks is recognized by Bitella et al. in [4] for the monitoring of soil water content. This change in system design allows for a broader range of development from independent researchers working towards the goals of sustainable agriculture at a reduced cost. The development of low cost, open source sensors and technologies for monitoring precision agriculture systems allows for expanded capabilities, increased reliability, and widespread use.

A large variety of data types can be utilized to help provide insight into vegetative health and potential production yield. Photosynthesis is a process by which all land plants take in water and carbon dioxide to create energy and oxygen. This energy is utilized for plant growth and ultimately for human food production. The more photosynthetic activity occurring within a plant, the more growth it can undergo. Water is fundamental to this process.

As a plant enters a water stressed state, its stomata begin to close. The stomata is one of the main barriers in the process of transpiration and closes in order to reduce excess water loss during a period of stress. Since the plant receives more radiation than it requires during these periods, it is forced to dissipate the energy as heat. This phenomenon was one of the first used to provide an estimate of the photosynthetic activity within plants. Using heat as an indicator for plant health is limited due to its influence on outside forces and changes to photosynthetic pigments in water stressed crops.

Narrowband spectral responses have been used as an indication of plant health based on the reflectance from plant canopies at various wavelengths. These techniques rely on the scattering and absorption mechanisms of surfaces from incident radiation and are less prone to error than thermal sensors. Often times the use of drones or other aerial monitoring devices are used to provide larger scale coverage outdoors. Recent studies have used spectral imaging techniques for the purposes of detecting viruses in plants by observing the polarized reflectance from their leaves [5], plant species discrimination [6][7][8], determining the relative water content of leaves using polarized reflectance [9] as well the general properties of leaf reflectance, transmittance and absorption[10][11]. Popular remote sensing vegetative indexes such as the Near Density Vegetation Index (NDVI) have been shown to be effective for determining the photosynthetic activity of vegetation be leveraging various spectral responses in the infrared and near infrared regions of light. Although spectral imaging techniques have higher degrees of accuracy, they are often cost prohibitive for smaller outdoor farms, and developing aerial monitoring mechanisms for large indoor agricultural operations is currently still being investigated. Application of low cost imaging devices for the monitoring of indoor agricultural operations allows for techniques in the visible and infrared range to be implemented at a larger scale. These devices can be mounted throughout a facility and positioned at a constant location or on a group of MAVs. Using Grey Level Co-Occurrence Matrices for determining the texture features of a given scene can easily be implemented using popular open source programming libraries such as scikit-image. This technique can therefore be implemented with any device that can capture a greyscale digital image and a computer running Python. GLCMs have been used in remote sensing for the purpose of classifying various types of terrain [12].

As reflectance models have developed from smooth ideal surfaces to complex multi-faceted bidirectional reflectance functions, polarization has also begun to be investigated as a property of these materials physiological and surface makeup. Light sources in different spectral ranges also play an important role in the polarization response of a material, but it can be shown that unpolarized input light can lead to a mathematically reduced polarization form, the polarizance response. This mathematical reduction simplifies data acquisition as only one linear polarizer is required to capture the details of a materials polarizance.

Previous techniques demonstrated in [13][14] involve the use of complex, expensive systems for calculating and performing a full analysis of the polarization properties of light as well as polarization response of materials. A simplified measurement scheme is presented here for a reduction in the overall number of required measurements, while still gathering important information about the materials' polarization response.

The consequent reduction in hardware cost and use of open source software for acquiring potential physiological indicators permits future infrastructure expansion using common technology stacks, such as Linux and Apache.

As the scarcity of fresh water increases, the need for preciously applying water as an input into crop production will also increase. Regions which generally have a lack of resources already, such as densely populated cities, have already begun investigating and implementing indoor agricultural production of crops. These types of controlled and monitored growing operations allow for the precise application of agricultural inputs and require large scale monitoring solutions.

The goal of this investigative study was to design and implement an efficient-low-cost polarization and texture based imaging technique for detecting the relative water content of vegetation.

1. Materials and Methods

The design of this experiment was intended to make the capture of polarization and texture based features simple and effective for modeling the relationship between extracted features and the relative water content of each sample. Individual leaves taken from an epipremnum aureum, or devils ivy plant as it is commonly called, were used as samples in this experiment. This particular species was used as it is commonly found and accessible. Overall thirty-four leaves were removed from their host and analyzed.

A digital microscope was used as a detector for measuring the light irradiated from the surface of each sample. This provided a larger field of view than traditional point detectors, and allowed for performing texture analysis on the surface properties of the leaves.

A single linear polarizer rotated to various positions in front of the detector was required to take the polarization measurements. From these images a simplified linear polarization response was calculated.

The images acquired by this detector were split into individual red, green, and blue color channels to analyze the potential spectral characteristics of the samples as well. Both polarization and texture features were extracted separately on each channel.

The extracted texture features used a Grey Level Co-Occurrence Matrix which describes the relationship between neighboring grey level pixel intensities in an image. Quantitive parameters can be extracted from this matrix to determine the texture features for a given sample through this technique.

Principal Component Analysis was then used for feature reduction and a linear regression was performed on the data. Statistical analysis was used to validate the model.

1.1. Imaging Techniques

Fresenels' equations dictate the behavior of transmitted and reflected electromagnetic waves from surfaces. It has been shown previously that at the Brewster angle all energy in the direction parallel to the plane of incidence is completely transmitted and the beam of reflected light is completely polarized in the perpendicular direction. This effect produces a high amount of polarization for smooth, ideal surfaces and is generally denoted as the specular portion of reflection. Diffuse reflection can be regarded as any reflection that is not specular. Previous studies have shown this portion of reflected light to be unpolarized [15], although more modern interpretations allow for polarization to be present in the diffuse portion of non ideal surfaces and useful for staging diseases[16] and surface orientation [11][17]. Since the specular component contains a majority of the polarization information, it was the primary focus of this investigation.

Two experimental setups were designed to capture the reflection of light from the leaves' surfaces in the specular and diffuse directions. The Brewster angle was determined to be approximately 55 degrees and was used for positioning the camera in the specular experiment. For the diffuse experiment the camera and polarizer were repositioned to be orthogonal to the plane of the sample, or 0 degrees. The experimental setup is shown for the specular detector orientation in Figure 1.

Surfaces which have more diffuse scattering are often considered to be rough when compared to smoother specular surfaces. Texture was therefore also considered by applying post processing software techniques.

A linear polarizer, digital microscope, and broadband light source formed the experimental setup. The broadband light source was found to have a low polarization, less than one percent, and was considered as an unpolarized input. An optical table was used to secure the components in place while the leaves were held in a vice, attached to a flat surface. Note the digital microscope was positioned as close as possible to the linear polarizer without being disturbed by the rotation of the polarizer.

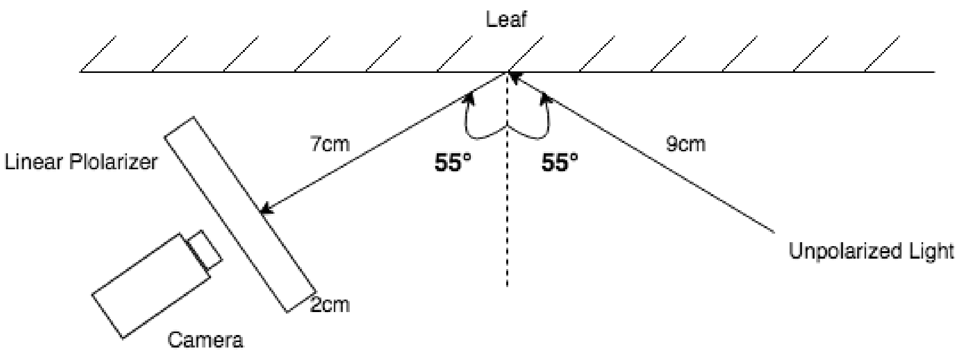

Figure 1. The experimental setup for capturing the specular reflection of light from each leaf sample's surface. A single polarizing lens is placed in front of a digital microscope to record the polarization measurements.

The polarization features were extracted by rotating the linear polarizer in front of the detector and images acquired were used for the polarization and texture analysis in this study. Each samples relative water content was determined independently using previously established techniques, described later.

1.1.1. RGB Image Analysis

A greyscale image is comprised of a multi-dimensional array of greylevel intensity values that range from 0 to 255. The shape of these arrays corresponds to the height and width of a given image in pixels.

Most digital cameras today are able to record color images as well. The color in images achieved by using a filter pattern arranged atop the image sensor which records the intensity of light through a filter sensitive to each of the primary colors, or spectral bands, in the red, green and blue regions of the visible light spectrum. Color images are stored as three multi-dimensional arrays containing the values of the red, green, and blue intensities for each pixel. Each color channel therefore can be represented as a greyscale image matrix of a pixel's filtered intensities and treated similarly during processing. Texture and polarization features were extracted independently from each of these individual channels.

1.1.2. Polarization Measurements

The polarization response of a material in its most general form is described by a 4x4 matrix known as the Mueller Matrix and is often denoted as M. Incident light of a known polarization state can be directed at the material in order to create a polarized output beam or response. This interaction is represented as

$$\mathbf{S}_{out} = \begin{bmatrix} S_{0out} \\ S_{1out} \\ S_{2out} \\ S_{3out} \end{bmatrix} = \mathbf{MS}_{in} = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} S_{0in} \\ S_{1in} \\ S_{2in} \\ S_{3in} \end{bmatrix} \tag{1}$$

The input $\mathbf{S}_{in}$ and output light $\mathbf{S}_{out}$ beam are formulated as 4x1 Stokes vectors which describe the overall amount of polarization the light beam contains as well as the relative strength of the polarization in orthogonal directions [13]. These directions are usually picked to be at 0, 45, 90 and 135 degrees for the linearly polarized portion of the light and in a right and left circular direction for the circular polarization properties. The circular components of the light beam are often left out of the discussion for simplification and data reduction [14]. Other studies have justified this reduction of dimensionality from the fact that most materials in nature have not been found to contain significant amounts of circular polarization [18].

In this reduced form, the dimension of the Mueller Matrix becomes 3x3 whereas the Stokes parameters are represented as 3x1 vectors. S_3 is therefore removed from the equations while S_1 describes the perpendicular and parallel components of polarization relative to the materials surface, and the S_2 component describes the polarization difference between 45 and 135 degrees. When unpolarized light is applied as the input vector

$$\mathbf{S}_{out} = \begin{bmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} m_{00} \\ m_{10} \\ m_{20} \end{bmatrix} \tag{2}$$

into this equation, the output polarization state is equal to the first column of the Mueller matrix. This column is said to be the polarizance of a material which is the property by which unpolarized input light becomes polarized by interaction with a sample [13]. The polarizance of a material is described

$$P(\mathbf{M}) = \frac{\sqrt{{m_{10}}^2 + {m_{20}}^2}}{m_{00}} \tag{3}$$

Although most outdoor skylight is partially polarized the controlled environment of the indoor experiment allowed for the use of a beam of light that was nearly completely unpolarized (<1 percent). The reduction in the Mueller Matrix created by utilizing unpolarized light as the input allowed for simplification in the experimental design for collecting the polarization properties of each sample. Capturing a full Mueller Matrix is a time intensive process that involves configuring polarization filters on the input light source as well as in front of the detector. Light measuring polarimeters can therefore be used to determine a partial polarization response of a material instead of needing a more complex Mueller matrix polarimeter. The polarizance property can be captured using a single linear polarizer in front of the detector which is rotated into four different positions. Under these conditions, these four measurements can be used to calculate the output Stokes parameters, which as shown, are the same as the first column of the Mueller matrix and therefore a measure of the sample's linear polarizance or reduced polarizance response.

These polarizance values can also be calculated as

$$\mathbf{P} = \begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} p_H + p_V \\ p_H - p_V \\ p_P - p_M \end{bmatrix} \frac{watts}{m^2} \tag{4}$$

where p_H, p_V, p_P and p_M represent flux measurements recorded through filters that extinguish orthogonal polarization states. This is a discrete polarimetric measurement and calculation.

These parameters can be normalized by dividing by the total intensity of the orthogonal images,

$$\frac{P_1}{P_0} = \frac{p_H - p_V}{p_H + p_V} \tag{5}$$

$$\frac{P_2}{P_0} = \frac{p_P - p_M}{p_P + p_M} \tag{6}$$

Note that $P_0 = p_H + p_V = p_P + p_M = p_R + p_L$ as each element represents the flux which passes through each orthogonal pair of linear polarizers. The normalized values are denoted P_1 and P_2 throughout the rest of this experiment.

Hence by recording the output polarization state of the unpolarized input through four linear polarizer positions, it is possible to capture information on the polarization response of the material. It is also important to note that a material must have polarizing properties for a polarizance response to occur.

Images were acquired for each sample with the digital microscope placed behind a linear polarizer and oriented at 0, 45, 90 and 135 degrees in order to acquire measurements for p_H, p_P, p_V and p_M. Each image was converted into a flattened array and the pixels with a value of 255 were removed from the dataset. At the highest intensity value, or the brightest points in the image, there is a risk of over saturation since the imaging device is limited to a maximum range. A recorded pixel value of 255 represents intensities that are equal to 255, or greater than this value since after this point, the sensor becomes saturated. This causes a skew in the data which is removed by the process of thresholding at this level. The same pixels were removed in each pair of orthogonal images. Due to areas of over saturation in the image, shadowing is a problem which also skews image data. In effect when certain pixels within an image are oversaturated, other pixels will be shadowed. Filtering pixels which were 0 for either image of the pair were therefore also removed to minimize this effect. These images were acquired in both the diffuse and specular direction for the reflected light. For each pixel in the red, green, and blue color channels, P_1 and P_2 parameters were calculated. The average polarizance values and standard deviation were calculated for all pixels in the image and added to the feature array for the sample.

1.1.3. Texture Analysis

The multidimensional pixel arrays of an image contain information for classifying the texture of a given scene. A Grey Level Co-Occurrence Matrix (GLCM) is a tool for classifying the texture of an image. By inspecting the grey level intensity for a given sample of adjacent pixels' relationships can be quantified as to the texture characteristics for that particular sample. These characteristics have previously been used for classifying various types of landscape [19][12] from overhead drone and satellite imagery. This quantitative measure of texture can be extracted from images and used as features of a dataset. The three groups of texture parameters which can be derived from a GLCM are contrast, statistical and orderliness.

A GLCM is able to quantify the spatial frequency distribution of grey level pixel intensity pairs for an image. A relationship between a reference pixel and neighbor is set a priori to determine the direction for grey level comparison. This angular relationship is chosen in multiples of either 0 or 45 degrees. Common GLCM spatial relationships are 0, 45, 90, and 135 degrees as shown in Figure 2.

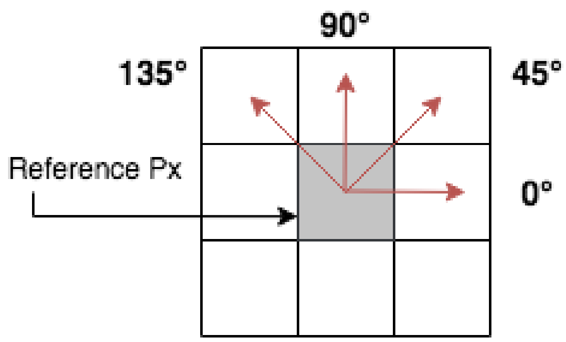

Figure 2. Each reference pixel is analyzed with respect to a given direction under inspection. For a given direction the grey level value of the pixel being pointed to is recorded. These directions are averaged out to remove spatial considerations for the texture quantification.

In order to quantify a texture in a rotationally consistent fashion, all four relationships are usually calculated and averaged together in determining the overall GLCM matrix. By measuring all four of these directionality's and averaging the GLCM features, the spatial directionality characteristic of the GLCM is removed and the texture is the same viewed from any direction. The GLCM has a size of NxN where N is the discrete quantized levels of the captured grey scale image. A single relationship Co-Occurrence matrix is formulated such that,

$$\mathbf{\Phi}_{ij}(\triangle x, \triangle y) = \sum_{x=0}^{n} \sum_{y=0}^{m} \begin{cases} 1, if I(x,y) = i \quad \text{and} \quad I(x+\triangle x, y+\triangle y) = j \\ 0, otherwise \end{cases} \tag{7}$$

where $I(x,y)$ is an nxm image and $\triangle x$, $\triangle y$ represent the predefined offset of the grey level pixel neighbor intensity relationship (i,j). Being defined as referencing one pixel to its neighbor to the right (0 degrees) the GLCM matrix is formulated as such, Non symmetrical GLCMs should be symmetrized by adding each to its transpose,

$$\mathbf{\Phi}' = \mathbf{\Phi} + \mathbf{\Phi}^T \tag{8}$$

Normalizing the frequency to one by dividing the matrix by the sum of all its elements results in a probability distribution for each grey level pixel pair.

$$\mathbf{P} = \frac{\mathbf{\Phi}'}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \mathbf{\Phi}'} \tag{9}$$

Features can then be extracted from the formed matrix for the purpose of defining single quantitative values for texture. These features are known as Haralick features and generally fall into 3 distinct feature categories; Contrast, Statistical and measures of Orderliness.

Contrast measures are defined by weights that increase or decrease with distance from the GLCM diagonal. These weights can be linear, exponential, etc. For the N x N dimensional GLCM matrix the N - 1 term in the first row or column represents pixel relationships that are of the greatest intensity difference.

Contrast, for example, has weights that increase exponentially away from the diagonal. It is calculated as

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P_{ij} \tag{10}$$

While dissimilarity is a measure of contrast with weights that increase linearly away from the diagonal

$$Diss = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i-j| P_{ij} \tag{11}$$

Statistical measures utilize each individual element of the GLCM as weights to determine the moments of the probability distribution matrix. No measures from this category were evaluated in this study.

Measures of orderliness are quantified by the amount of entropy and energy within an image. Entropy is a measure of randomness in a system. In thermodynamics, it is the recorded heat lost when a reaction occurs; a measure of disorder. Energy is a measure of useful work that can occur due to the nonrandom nature of the energy in a system. So for an image, higher randomness in the grey level tones of neighboring pixels results in a higher amount of entropy for the GLCM matrix associated with the image. The angular second moment (ASM) describes the amount of "inertia" around a pixel neighbor relationship and is defined as,

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij}^2 \tag{12}$$

The square root of the ASM results in the energy of the system

$$Energy = \sqrt{ASM} \tag{13}$$

For perfectly uniform textures the energy will be at a maximum of 1 [20]. This experiment captured the energy and correlation features of each individual red, green and blue image channel by averaging the GLCM over 0, 45, 90, and 135 degree relationships. Twenty image samples, each measuring 75 by 75 pixels, were extracted from each samples H aligned polarization filter. Texture was extracted from each color channel. The average of each samples texture measure were calculated and added to the feature array for each sample. Further measures could also be extracted for further analysis, although it is recommended that texture features are selected from each of the three categories mentioned.

1.2. Relative Water Content

A variety of parameters have been investigated to understand the physiological condition of plants such as relative water content, water stress, chlorophyll content, etc. Calculating the relative water content is a process that is time consuming and slow. It involves a destructive measurement of each leaf by removing it from the plant and performing a series of weight measurements. These measurements involve taking the freshly cut weight of the leaf, a turgid weight, and a dry weight. The general procedure for determining the relative water content of individual plant leaves has been discussed in [21] and in brief summary is as follows

1. Remove leaf from host plant leaving approximately 2 cm of petiole
2. Weigh leaf to acquire the Fresh Leaf Weight (FW)
3. Place leaf petiole in solution of distilled water and $CaCl_2$ at 2mM for at least 8 hours
4. Weigh leaf to acquire Turgid Weight (TW)
5. Place leaf in an oven at 60°C for 4 days
6. Weigh leaf to acquire the Dry Weight (DW)

The relative water content can then be calculated as a percentage,

$$RWC = \frac{FW - DW}{TW - DW} x100 \tag{14}$$

Note that the scale used for weighing needs to have at least 4 decimal places to ensure the accuracy of the measurements. Drying times and artificial hydration times can vary with species and oven temperature. The process of acquiring the relative water content of leaves is destructive and requires several days to get the required measurements. A goal of this experiment was to create a simple effective measurement for determining the RWC of a plant without destructively removing its leaves and reducing the time it takes to get experimental results. The features extracted from the images were investigated and modeled against the RWC measurements acquired. It has been noted that a plant's physiology experiences changes in relation to the amount of RWC a plant has. It is noted that

This study mainly focused on leaves which were in the 90 - 100 % range and were experiencing the closing of their stomata as described in Table 1. "An increase in reflectance. . . is not directly related

Relative Water Content (%)	Plant Physiological Response
90–100	closing of the stomata, reduction of cellular expansion and growth
80–90	tissue composition change, altered rates of photosynthesis and respiration
<80	ceasing of photosynthesis

Table 1. Plant physiological responses to detected relative water content levels.

to water content but indirectly, since a decrease in water content can lead to an increase in internal leaf air space or cell breakdown which may increase reflectance and decrease transmittance [22]".

This increase in internal air space leads to multiple scattering at air wax boundaries, and creates differences in the reflection, transmission and absorption of light, and the P_1 and P_2 polarizance parameters of the response.

Field measurements of the physiological properties of plants are time consuming and error prone. It is therefore beneficial to pursue solutions that quantify these metrics in large area field measurements.

1.2.1. Principal Component Analysis

When dealing with high dimensionality datasets, it is important to asses the correlation between each of the features so as to not have duplicate information. As datasets also grow larger it becomes more difficult to visualize the data and experimental results, and the time it takes to compute the experimental results increases. Principal Component Analysis is a technique which aims to reduce the dimensionality of a given dataset while maintaining the characteristics of each feature that provide the least amount of correlation and the highest amount of explained variance. Prior to performing PCA, features are usually normalized to have a variance of one and a mean of zero allowing for features on different scales to be viewed equally during the model development. After normalization, eigenvalue decomposition is performed on each of the feature sets to maximize the variance of each principal component. 12 features were extracted from each sample's images, normalized, and reduced into two principal components. This reduction allowed for ease of analysis in three dimensional space.

1.2.2. Linear Regression

A multivariate ordinary least squares regression was performed on the two principal components and their samples corresponding measured RWC. This led to an understanding of the relationship between the acquired feature vectors of each sample and how they relate to the relative water content.

Linear regression analysis has long been in use in the field of statistical and supervised learning. They provide the ability to predict quantitative responses, Y, when X is a set of inputs. The assumption when using this technique is that the relationship between these two variables is linear, and of general form

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X_1} + \beta_2 \mathbf{X_2} \tag{15}$$

where β_0, β_1, and β_2 are parameters that are calculated using a set of input data and represent the intercept and slope of the regression. Once trained, this model can predict future output values for a given input. When given a set of observations, β_0, β_1, and β_2 are calculated in order to have a closeness between the predicted line and the observed data. A common measure of this closeness is the least squares error. The residual, e, for a given set of observations is calculated as

$$e_i = y_i - \hat{y}_i \tag{16}$$

where $\hat{y}_i$ are the predicted outputs. These residuals can be used to calculate the Residual Sum of Squares RSS, or the amount of variation left unexplained after performing the regression. It is

$$RSS = {e_1}^2 + {e_2}^2 + \cdots + {e_n}^2 \tag{17}$$

where n is the number of observations in a dataset.

The Total Sum of Squares is the measure of how much variability exists within the data before the regression has been performed. It is defined as

$$TSS = \sum_{i=0}^{n} (y_i - \bar{y})^2 \tag{18}$$

where $\bar{y}$ is the mean. Using the RSS and TSS allows for determining the accuracy within the model by calculating the R^2, or the "proportion of variability in Y that can be explained by X". It is defined as

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \tag{19}$$

For a given application it is difficult to determine what a 'good' R^2 score is, although it has been noted that in biological applications there can be a great deal of unexplained variance, sometimes even less than 0.1 [23]. In these experiments, linear regression was utilized to correlate the relative water content of the devils Ivy plant leaves with the first and second principal components derived from texture and polarization characteristics. R^2 was used to gauge the accuracy of the results along with other statistical measures.

2. Results

The feature extraction was performed using Python v3.6.4, OpenCV2, scikit-image v0.14.2, and scikit-learn v0.20.3 packages. The raw images acquired during the experiment were placed into a sample directory along with an rwc.dat file which contained the relative water content for the sample. As these features and data were extracted for each sample they were written to a csv file for storage and future analysis. As more features are extracted from each sample, the processing power required by a computer increases. Writing the dataset to a csv file saves on processing time when analyzing the data which were then analyzed and plotted using the pandas, statsmodel and seaborn python packages.

Prior to performing PCA on the feature set to reduce the overall dimensionality, a correlation matrix was used to visualize the correlation between each feature. The correlation matrix is shown in Figure 3. In the ideal case, uncorrelated features are used to reduce the redundant information in the dataset. The diagonal of a correlation matrix represents the amount of correlation each feature has with itself, and is always one. All other feature combinations are shown to be either positively are negatively related. Features from the texture category are shown to be highly correlated to one another.

For the specular H image acquired, the RGB channel analysis had little impact on texture as the various color channels showed the same amount of correlation for similar parameters under consideration.

This representation allows for insight into how separate RGB channel analysis performs in that texture characteristics are not as affected by color separation for the H filtered image, although in the V filtered image there is more of a distinction. This aligns with previous reasoning that the V filtered image should include the least amount of polarization as it is filtering it out. This is due to there being less white light in the image and the quantified texture is derived solely from the leaves' surface.

Similarly polarization features were more highly correlated with one another, although RGB channel analysis shows more of a distinction between the various polarization features for each

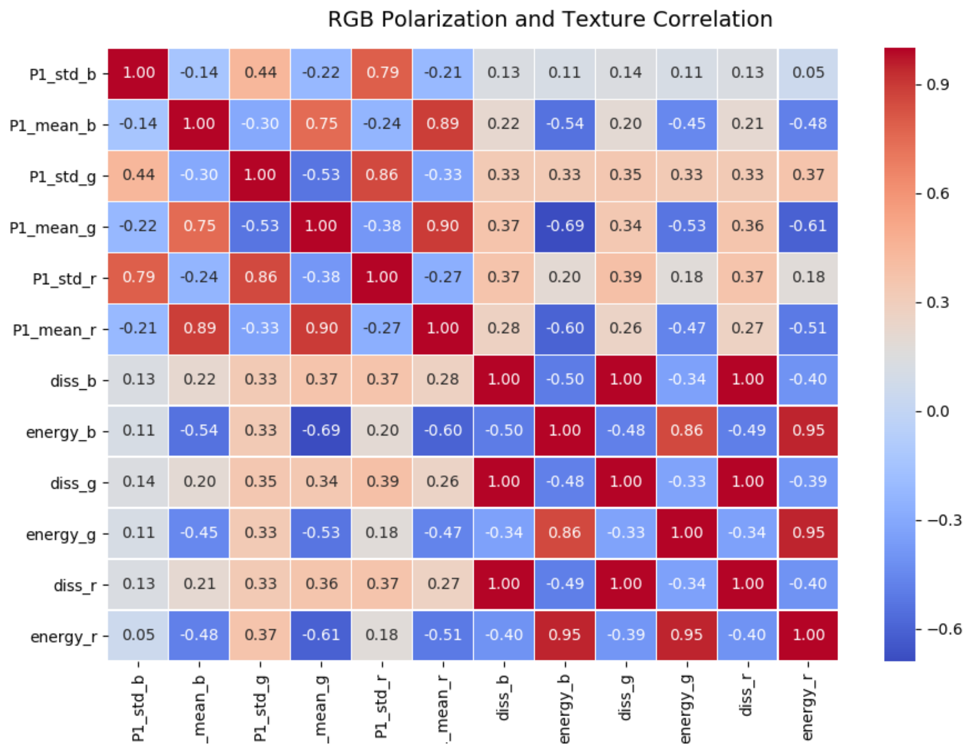

Figure 3. A correlation matrix showing the relationship between each feature accross the input feature set. Note that texture and polarization features are the least correlated, while RGB channel analysis appears to have no effect on the texture analysis.

channel. This agrees with the premise that polarization is a frequency dependent phenomenon. Polarization filters are typically specified for a given frequency response range or broadband response, while equations generally are designed for single frequency properties. Using the individual color channels of the image sensor allows for a more bandlimited response when compared to its grey level counterpart. Correlation between texture and polarization filters showed the lowest value and therefore provide a good basis for principal component creation.

The use of features from each of these categories provided insight into the light interaction at the surface of the leaves and the different results produced from each feature set. Principal component analysis was performed on each component resulting in two principal components made of of each feature and weighted such that the variance explained of each feature for a given component was maximized.

These two principal components accounted for 74.75 percent of the total variance. A linear regression was fit to these two principal components and a linear regression performed for each samples' measured relative water content.

Due to the reduction of feature data into two components, the data could be more easily visualized on a three dimensional graph. The two principle components were graphed in the x,y plane and the z axis was assigned to the RWC value of each sample with these values plotted as a scatter plot. A mesh grid was created to cover a range of inputs for the samples that ranged from the lowest to highest value of each component. This array of values was then input to the fit linear regression equation and plotted as a plane. The resulting regression is shown in Figure 4 as a plane plotted against each PC.

The results of the statistical analysis for this regression showed an R^2 of 0.409 and an adjusted R^2 of 0.371. An adjusted R^2 score is used when there are multiple features as input to the regression and accounts for the arbitrary increase in R^2 as more features are added to the model. The adjusted R^2 is penalized for additional features added.

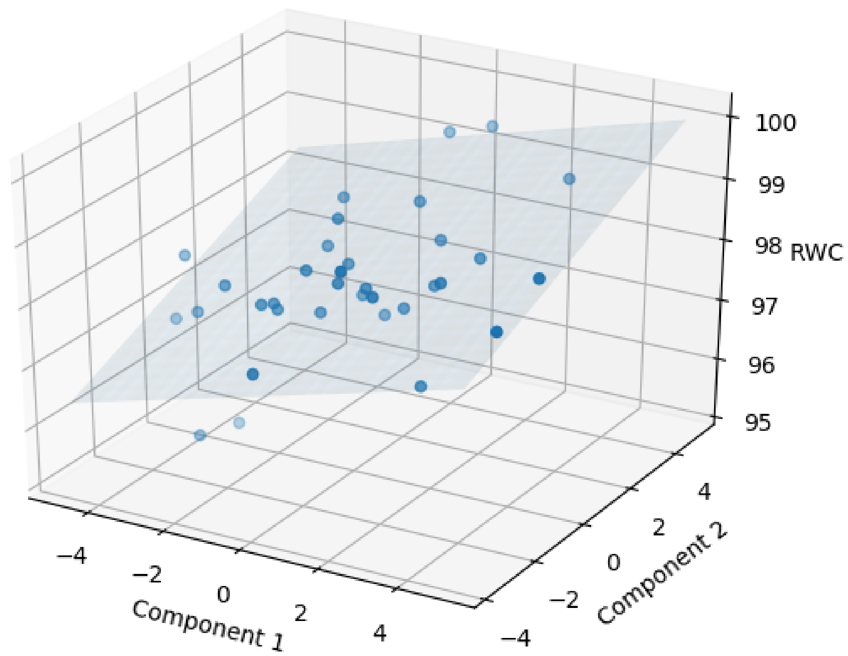

Figure 4. A 3d representation of the relationship between the two principal components and the relative water content of each sample. The solution plane is shown for the calculated multiple linear regression.

Dep. Variable:	RWC	**R-squared:**	0.409
Model:	OLS	**Adj. R-squared:**	0.371
Method:	Least Squares	**F-statistic:**	10.75
Date:	Tue, 09 Apr 2019	**Prob (F-statistic):**	0.000285
Time:	21:35:17	**Log-Likelihood:**	-36.903
No. Observations:	34	**AIC:**	79.81
Df Residuals:	31	**BIC:**	84.39
Df Model:	2		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	98.0273	0.129	761.878	0.000	97.765	98.290
PC1	0.2170	0.055	3.926	0.000	0.104	0.330
PC2	0.1684	0.068	2.466	0.019	0.029	0.308

Table 2. Statistical analysis results for the linear model.

The statistical analysis of the data and regression can be found in Table 2. The parameters for the linear equation that result can be shown as the solution to a plane such that

$$\mathbf{Y} = 98.0273 + 0.2170\mathbf{X_1} + 0.1684\mathbf{X_2} \tag{20}$$

where X_1 is $PC1$ and X_2 is $PC2$.

The F statistic relates the mean sum of squares to the mean error sum of squares. It is a test of the regression model under the null hypothesis. A low F statistic probability shows the probability of the parameters of the model being zero is low and the regression equation is valid for fitting the model. This means that the models' independent variables are not purely random with respect to the dependent variable. The constant coefficient shows what the y intercept would be if both PC1 and PC2, were zero. This model could benefit from additional samples and a more varied RWC, but due to experimental limitations, samples at lower RWC were not available. For each parameter the STD error is low and shows that each coefficient has a high level of accuracy. The P value is a common statistical measure that asserts how much confidence there can be in the results. Generally a P value of less than 0.05 is considered statistically accurate and there the measurements of this experiment can also be concluded to be statistically accurate.

These results show that there is a relationship between the relative water content of devils ivy leaves with the polarization and texture response captured by images during the course of the experiment. Further investigation is needed into a larger range of RWC measurements to validate further this preliminary study. It has been shown that with consumer grade electronics, it is possible to derive features from organic samples for processing and analysis. As more data is acquired, more advanced models could be developed. Data for other species of vegetation should also be investigated for varying curves.

3. Discussion

Although high precision narrowband spectral sensors are still cost prohibitive for many operations, low cost imaging devices can be useful for agricultural monitoring applications. As this study shows, it is possible to use polarization and texture based imaging techniques for extracting health indicators from vegetation.

As plants experience water stress their physiological and structural makeup change. These changes affect the scattering of incident radiation and have been shown to provide insight into the overall health of the plant. Previously, spectral signatures have been used to determine the RWC of vegetation [24] and determining the water status of canopies continues to be a goal of remote sensing research. As these studies have progressed and produced results, the implementation of these ideas and technologies on other areas of agricultural production has grown. Aerial sensors attached to Micro Aerial Vehicles (MAVs) [1] have been implemented to collect data for monitoring greenhouse production of crops for reducing the overall cost of these systems. Open source hardware has been used to build sensor networks for collecting soil moisture content at a low cost which has lowered the barrier for implementing these technologies at different scales.

The polarization response of plants has extended models that previously focused solely on the reflected irradiance from surfaces [25]. It was previously shown that the polarization of the diffuse portion of irradiance from a leafs' surface provides less information than the specular portion of reflection [15]. Although the diffuse portion of polarization have been shown to be useful for disease staging in other areas, this study shows that more information is held in the specular portion of polarization. The sensitivity of the imaging device also has an impact on the ability to detect polarization in the diffuse portion of reflectance.

Many of these studies require the use of lasers at specific wavelengths for measuring and calculating a vegetative index. Use of this type of radiation requires multiple lenses for delivery and measurement. Application of these techniques are cost prohibitive. The experimental setup described here provides an overall reduction to the image processing chain and results in a simplified polarizance response which can be measured using a single lens.

Textual studies from satellites have been used for determining weak portions of levees and assessing environmental impact scenarios [12]. These texture measurements are useful as they can be calculated after the image has been captured. This allows for the image acquisition phase to remain as simple as possible, while maximizing the features extracted from a given scene. It was shown that the texture features and polarization features acquired in this study are weakly correlated, and therefore serve as good candidates for principal component analysis as they do not contain redundant information. A combination of polarization and texture features provided the best results.

Due to the effect of growth stage on leaf scattering a more controlled experiment would entail recording the growth stage of each plant while performing this study. An expanded dataset should then be acquired to address the RWC of various species across a wider range of water stressed states. Water application would be closely monitored and controlled throughout the experiment. With this expanded dataset, more generalized models could then be established for practical application. Additional sensor types could be experimented with as well as the collection of additional physiological indicators used to determine plant health.

The use of open source software allows for these experimental results to be reproduced, shared, and expanded upon using a set of community standards. The portability of Python makes it ideal for experimental use as most computers today have Python installed. All original samples and code are available on Github. This combined with cost effective hardware allows for scalable and robust agricultural monitoring capabilities for the future.

As more investigation is performed across different species, large scale implementation of these sensors indoors becomes feasible for the precise application of water. This will be increasingly important as fresh water becomes scarce. Additional physiological properties such as chlorophyll composition and growth stage could also be investigated for accurately monitoring agricultural plant health, thereby promoting the long term goal of reducing agricultural inputs while maximizing outputs in precision agricultural.

4. Conclusion

During the course of this study an efficient-low-cost sensor for determining the relative water content of vegetation was designed, implemented, and tested. The design provided a simplified image acquisition approach to reduce the number of hardware components needed and reduce cost. Post processing of the acquired images from the experiment produced texture and polarization features for each individual leaf sample. These features were linearly correlated with the relative water content of the individual samples.

The use of open source software in this study allows for contribution by other researchers interested in development of precision agriculture and open data access in this field. Future experiments should utilize open source hardware to better understand the imaging sensor design and characteristics, a more controlled growing environment to more closely monitor the effects of water stress, different physiological indicators such as chlorophyll, fully automated polarization acquisition for improved efficiency, and a multi-sensor node deployment.

Collecting datasets of physiological indicators and imaging data for a variety of vegetative species will be needed to provide large scale insight into the health of ecosystems and advance the long term goals of precision agriculture. Natural resources are becoming increasingly scarce, and the need for minimizing the inputs to agricultural systems while maximizing their output, while monitoring the overall state of the Earth's ecosystem, will continue to be of importance.

Author Contributions: Nicholas Ericksen performed the conceptualization of precision agriculture concepts related to imaging techniques. Writing of software for image processing and data analysis. Writing of the original draft.

Author Contributions: Romeo Pascone provided supervision of laboratory experiments. Research of polarization and texture methodologies. Formal analysis of proofs.

Funding: This research received no external funding

Acknowledgments: Computational support and experimental assistance from the Departments of Mathematics and Biology of Manhattan College

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Longhi, M.; Taylor, Z.; Popovic, M.; Nieto, J.; Marrocco, G.; Siegwart, R. RFID-Based Localization for Greenhouses Monitoring Using MAVs. *2018 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)* **2018**. doi:10.1109/APWC.2018.8503764.
2. Panda, G. Saha, T. Building of a Low Cost Reliable Wireless Sensor Network for Smart Indoor Agriculture Products. *2nd International Conference on Electronics, Materials Engineering, and Nano Technology* **2018**. doi:https://doi.org/10.1109/IEMENTECH.2018.8465206.
3. Cao-hoang, T.; Duy, C. Environment monitoring system for agricultural application based on wireless sensor network. *Seventh international conference on information science and technology* **2017**. doi:https://doi.org/10.1109/ICIST.2017.7926499.
4. Bitella, G.; Rossi, R.; Bochicchio, R.; Perniola, M.; Amato, M. A Novel Low-Cost Open-Hardware Platform for Monitoring Soil Water Content and Multiple Soil-Air-Vegetation Parameters. *Sensors* **2014**, *14*. doi:https://doi.org/10.3390/s141019639.
5. Maxwell, D.; Partridge, J.; Roberts, N.; Boonham, N.; Foster, G. The Effects of Plant Virus Infection on Polarization Reflection from Leaves. *PLoS ONE* **2016**, *11*. doi:https://doi.org/10.1371/journal.pone.0152836.
6. Bosch, R. Plant Classification System for Crop / Weed Discrimination Without Segmentation. *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference* **2014**. doi:https://doi.org/10.1109/WACV.2014.6835733.
7. Bhardwaj, A.; Kaur, M.; Kumar, A. Recognition of plants by Leaf Image using Moment Invariant and Texture Analysis. *International Journal of Innovation and Applied Studies* **2013**, *3*, 237–248.
8. Askraba, S.; Paap, A.; Alameh, K. Laser-Stabilized Real-Time Plant Discrimination Sensor for Precision Agriculture. *IEEE Sensors Journal* **2016**, *16*. doi:http://dx.doi.org/10.1109/JSEN.2016.2582908.
9. Vanderbilt, V.; Grant, L.; Daughtry, C. Polarization of Light Scattered by Vegetation. *Proceedings of the IEEE* **1985**, *73*. doi:http://dx.doi.org/10.1109/PROC.1985.13232.
10. Bousquet.; Lacherade.; Jacquemoud.; Moya. Leaf BRDF measurements and model for specular and diffuse components differentiation. *Remote Sensing of Environment* **2005**, *98*, 201–211. doi:https://doi.org/10.1016/j.rse.2005.07.005.
11. Atkinson, G.; Hancock, E. Recovery of Surface Orientation from Diffuse Polarization. *IEEE Transactions on Image Processing* **2006**, *15*. doi:http://dx.doi.org/10.1109/TIP.2006.871114.

12. Dabbiru.; Aanstoos.; Ball.; Younan. Screening Mississippi River Levees Using Texture-Based and Polarimetric-Based Features from Synthetic Aperture Radar Data. *Electronics* **2017**.
13. Chipman, R. *Polarimetry*; Handbook of Optics, McGraw-Hill Education, 2009.
14. Goldstein, D. *Polarized Light*; CRC Press, 2003.
15. Grant, L. Diffuse and Specular Characteristics of Leaf Reflectance. *Remote Sensing of Environment* **1987**, *22*, 309–322. doi:https://doi.org/10.1016/0034-4257(87)90064-2.
16. Giakos, G. Polarimetric Phenomenology of Light Interaction with Early Lung Cancer Stages. *Measurement Science and Technology* **2011**, *22*.
17. Atkinson, G.; Hancock, E. Shape from Diffuse Polarisation. *Proceedings of the British Machine Conference* **2004**. doi:http://dx.doi.org/10.5244/C.18.94.
18. Marshall, J.; Cronin, T. Patterns and Properties of Polarized Light in Air and Water. *Philos Trans R Soc Lond B Biol Sci* **2011**, *1565*. doi:http://dx.doi.org/10.1098/rstb.2010.0201.
19. Soh, L. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. *IEEE Transactions on Geoscience and Remote Sensing* **1999**, *37*, 780–795. doi:https://doi.org/10.1109/36.752194.
20. Haralick, R. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* **1973**, *SMC*, 610–621. doi:https://doi.org/10.1109/TSMC.1973.4309314.
21. Reigosa Roger, M. *Handbook of Plant Ecophysiology Techniques*; Springer Science & Business Media, 2001.
22. Myneni, R.; Ross, J. *Photon Vegetation Interactions*; Applications in Optical Remote Sensing and Plant Ecology, Springer-Verlag, 1991.
23. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*; Springer Texts in Statistics, Springer, 2017.
24. Ihuoma, S.; Madramootoo, C. Crop reflectance indicies for mapping water stress in greenhouse grown bell pepper. *Agricultural Water Management* **2019**. doi:10.1016/j.agwat.2019.04.001.
25. Meister, G.; Wiemker, R.; Monno, R. Investigation on the Torrance-Sparrow specular BRDF model. *Geoscience and Remote Sensing Symposium Proceedings* **1998**. doi:http://dx.doi.org/10.1109/IGARSS.1998.703752.

Sample Availability: All raw data and code for calculations can be accessed on https://github.com/nicholasericksen/glcm_polarization_rwc_devils_ivy/tree/master