Classification of Tomato Growth Degree Adopting Machine-Learning to Photomorphogenesis Information in the Visible Light Region

1st Yoshitsugu Nakagawa
Electronics Technology Group, Tama-Techno Plaza
Tokyo Metropolitan Industrial Technology Research Institute
Tokyo, Japan
nakagawa.yoshitsugu@iri-tokyo.jp

2nd Hiroyasu Sano
Electronics Technology Group, Tama-Techno Plaza
Tokyo Metropolitan Industrial Technology Research Institute
Tokyo, Japan
sano.hiroyasu@iri-tokyo.jp

3rd Toshiro Takata
Nozomi Corporation.
Tokyo, Japan
takata@nozomicorp.jp

***Abstract*— In recent years, the introduction of IoT has advanced across all industrial sectors. Efforts toward "smart agriculture" aim to enhance agricultural management efficiency by visualizing plant cultivation, improving crop quality, and collecting various data regarding crop growth to understand and analyze trends. This study focuses on the light spectrum necessary for crops to grow under solar radiation and extracts wavelengths within the visible light range (400–700 nm) from scattered light with a broad spectrum. Multi-spectral sensors can measure the reflectance of light at each wavelength by passing it through silicon filter lenses, capturing the "distribution of crop growth" as information obtained from light. In this paper, we sample the fruit portion of tomatoes as the target crop and discuss the changes in distribution between wavelengths that are not visible to the naked eye, based on differences in light reflectance. Tomatoes change color during the growth process through photosynthesis, eventually accumulating acidity and sweetness, reaching the harvest time. At this point, the distribution between wavelengths can be quantified, allowing us to interpret the optimal harvest timing. While farmers have traditionally relied on their experience to judge fruit ripeness based on color, using this quantified indicator is expected to greatly assist new farmers entering agricultural work. Furthermore, by accumulating these quantified datasets, it will be possible to apply them to machine learning for classifying the optimal harvest timing for future data samples. This paper also presents the methodology for this process.**

***Keywords*— light multispectral, photomorphogenesis information, crop growth degree, visible light, smart agriculture**

I. I. INTRODUCTION

In the field of agriculture, the implementation of various technologies, including IT utilization, drones, automated guided vehicles, and plant factories, continues to contribute to the increased productivity of relatively large-scale agricultural operators. In particular, research and field trials regarding agricultural IT in crop growth are diverse. The Ubiquitous Environmental Control System (UECS) technology enables remote monitoring of temperature and humidity control, sunlight restriction, irrigation electromagnetic valve control, soil component management, etc., in facility cultivation that can be isolated from external environmental changes and pests. This eliminates the need for physical presence and visual assessments, achieving multifunctional and automated environmental control. Additionally, techniques have been reported that use hyperspectral cameras mounted on drones to analyze the color distribution of crops growing in fields, allowing for the determination of growth trends and pest damage [1][2]. By using specific wavelengths of light emitted by LEDs, which promote growth as a substitute for sunlight, high-density growing spaces can be created, enhancing value by constructing plant factories within the crop consumption area, thus improving product freshness and reducing transportation costs [3][4].

On the other hand, it is also true that there are many small-scale farmers who hesitate to implement these systems, aside from those agricultural businesses that can cover the costs of UECS installation and the rising energy consumption costs associated with environmental control. Especially in urban agriculture, where the arable land area is limited in urban areas and their surroundings, there is a need to focus on high-quality and high-value cultivation techniques that enhance sweetness and color, in addition to increasing yields to stabilize agricultural management. Moreover, the decline in the farming population and the aging of farmers are seen as an impending crisis that threatens crop self-sufficiency, and it has become a responsibility to delve into and research agricultural IT, including the acquisition of new farmers and skill inheritance.

Traditionally, agricultural workers engaged in farming while judging the growth of crops and environmental conditions based on their own experiences, know-how, and subjective sensibilities, conveying this knowledge to successors over time [5][6]. To address this issue, this research aims to establish a technology that visualizes and substantiates the conditions implied by subjective sensibilities by quantifying the degree to which crops grow through nutrition and photosynthesis using light spectra. Furthermore, it seeks to accumulate and analyze data to estimate the growth process of crops and aid in future predictions.

II. PROTOTYPE OF EQUIPMENT FOR MEASURING SPECTRUM

A. *Scattered Light and Visible Light Region*

As a general principle, the weight, nutritional components, and texture of crops are determined by photosynthesis, which produces carbohydrates and oxygen from carbon dioxide through the conversion of specific wavelengths of light energy into the chemical energy necessary for plant growth. The wavelengths of scattered light range from 300 to 3000 nm, but the wavelengths effective for photosynthesis are considered to be in the visible light range (400 to 700 nm). Traditionally, it has been necessary to use polarized filter lenses on optical sensors to extract specific wavelengths. In this research, we developed a prototype that uses a silicon-based visible light band filter in an array shape to measure the multispectral data of the visible light range through digital processing.

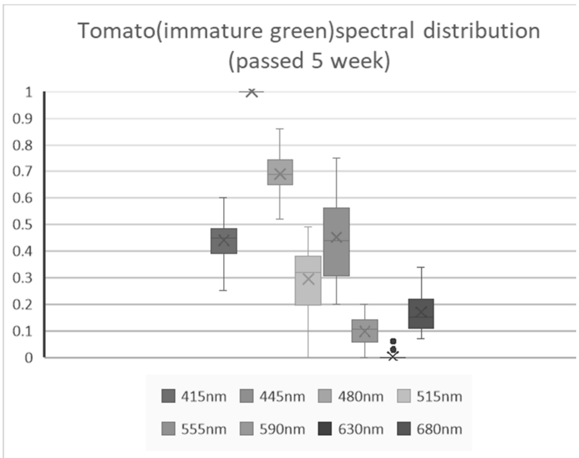

Figure 2. Show Spectral Distribution of Tomato Growth of immature green process as Boxplot.

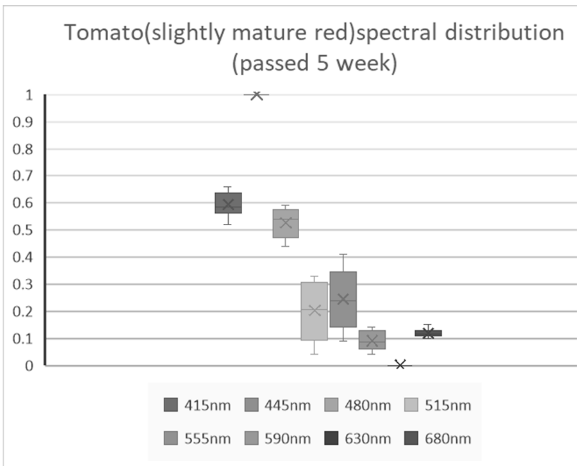

Figure 1. Show Spectral Distribution of Tomato Growth of slitely mature red process as Boxplot.

TABLE I. PARAMETERS FOR EXPOSURE CONTROL.

Step	Gain	Time
1	0.5x	140ms
2	2x	140ms
3	8x	140ms
4	32x	140ms
5	128x	140ms
6	512x	140ms
7	512x	140x 4ms

Note: When the maximum light intensity value satisfies conditional expressions (2) and (3), the exposure condition is updated from outdoors (Step 1) to a darkroom (Step 7).

B. *Light Morphogenesis Information*

Light morphogenesis information refers to capturing not only visible colors but also the role of information obtained from light. By extracting specific wavelength components from the broad spectrum of scattered light, we observe the degree of light reflection of crops to capture their characteristics and monitor changes occurring during the growth process. The growth rate of crops is believed to vary under different conditions, such as the locations where fruits appear and the density of leaves. Furthermore, even if the same color is perceived by the human eye, it becomes possible to identify the progress of growth by contrasting the obtained information. The prototype captures eight wavelengths of light (415, 445, 480, 515, 555, 590, 630, 680 nm) in the visible light range as a light multi-spectrum (Figure1,2).

C. *Normalization Processing of Light Wavelengths*

The amount of sunlight received from the sun changes due to seasonal variations and weather conditions. When capturing light intensity, measurements that can adapt to these changes are required. The optical semiconductor built into the prototype controls exposure based on measurement time and aperture size to convert light reflection levels. If strong light intensity is received for a certain period while the aperture remains large, the light reflection level may saturate. Conversely, in indoor environments where sunlight does not directly enter, it is necessary to reduce the aperture size to increase sensitivity. In facility cultivation, the light intensity of scattered light also changes over time, so if a certain time and aperture size are maintained, it is possible that one of the light wavelengths may reach a saturated or low-resolution state. Additionally, when exposure conditions are changed, the magnitude (absolute value) of the light reflection level measured under the same light wavelength can vary depending on the exposure conditions. Therefore, normalization processing was performed to compare the relative values between light wavelengths. The formula related to the normalization processing is expressed as (1).

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Additionally, we established an algorithm for automatic exposure control that follows the rules of conditions (2) and (3). This processing confirmed that even when the exposure conditions change due to differences in time, there are no discrepancies in the relationship of distributions between light wavelengths. Here, represents the minimum light wavelength. The parameter variables for exposure control are shown in Table 1.

$$\begin{aligned} x_{max} &> \rho K \quad \rightarrow Step\ Up \\ x_{max} &< (1-\rho)K \rightarrow Step\ Down \end{aligned} \quad (2)$$

$$\begin{aligned} K &= V_{max} \\ &= 50000 \\ \rho &= 0.8 \end{aligned} \quad (3)$$

III. DATA PROCESSING TECHNIQUES FOR MACHINE LEARNING

In this section, we consider a machine learning classification model to distinguish the difference in the growth stage of tomatoes from the difference in wavelength distribution in optical sensing and describe the classification procedure.

A. Machine Learning

In recent years, the implementation of artificial intelligence (AI) in various contexts has enabled automatic learning from past data sample sets without human intervention, allowing for fact-based reasoning and classification to make judgments or decisions on behalf of humans. In other words, machine learning is a mechanism by which computer programs learn automatically to create mathematical models.

In this study, we aim to construct a classification model for a group of data representing the distribution of wavelengths measured and normalized by light spectrum, which is plotted as multiple classes of point clouds on a two-dimensional plane. The method used to create this model is the Support Vector Machine (SVM). Geometrically, it is formulated by maximizing the non-interference zone (margin) M, which separates each point cloud from the boundary line that classifies the point clouds (equation). Furthermore, the data used in this study consists of sampled data, which is treated separately as training data and test data. The point x plotted on a two-dimensional plane and the parameters W and b are formulated in vector form (4) as follows:

For vectors

$$\begin{aligned} \boldsymbol{x}_i &= (x_{i1}, x_{i2})^T \\ \boldsymbol{w} &= (w_{i1}, w_{i2})^T \end{aligned} \quad (4)$$

$$W^T X_i + b = 0 \quad (i = 0,1,2,\cdots,N) \quad (5)$$

The separating hyperplane in n-dimensional space is expressed by equation (5).

Next, this paper applies a classification model that allows

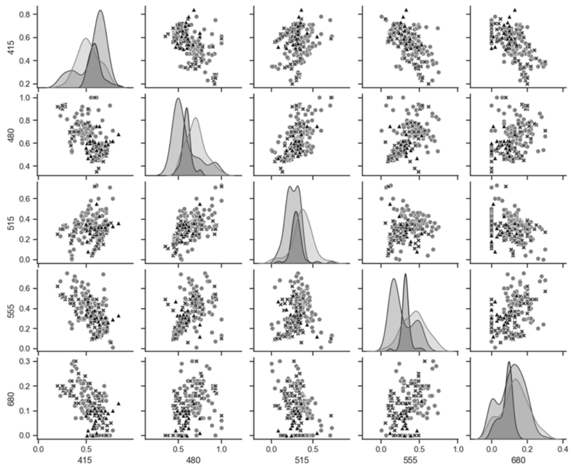

Figure 3. Show Spectral Distribution of Selected btween Light Wavelength in Visible light regin as Pairplot.

Note: The units for both the vertical and horizontal axes are [nm].

$$\max_{w,b} M, \frac{t_i(W^T X_i + b)}{\|W\|} \gg M \ (i = 0,1,2,\cdots,N) \quad (6)$$

$$t_i(W^T X_i + b) \geq 1 - \xi_i \xi_i = \max\left\{0, M - \frac{t_i(W^T X_i + b)}{\|W\|}\right\} \quad (7)$$

ξ_i units inside the margin (equation). This means that it also permits points to exist in the opposite region of the classification boundary. This is because there is a certain degree of variation in the fruit measurements sampled during the crop growth process. This variation may arise from measurement errors of the sensors in the actual measurements or from light scattering coming from different directions contaminating the sensor readings.

Here, ξ_i represents the degree to which the training data

x_i is allowed to protrude inside the margin, serving as a parameter when optimizing the SVM, represented by the following equation:

$$\min_{W,\xi}\left\{\frac{1}{2}\|W\|^2 + C\sum_i^N \xi_i\right\}$$
$$t_i(W^T X_i + b) \geq 1 - \xi_i \xi_i = \max\left\{0, M - \frac{t_i(W^T X_i + b)}{\|W\|}\right\} \quad (8)$$

The second term of the equation includes the objective function. A smaller cost parameter C allows ξ_i to be larger, resulting in looser constraints. Conversely, a larger C restricts ξ_i from being large, preventing the training data from existing inside the margin or crossing the classification boundary into the opposite region. C becomes a hyperparameter that determines the performance of the SVM.

B. Removal of Outliers

Verifying whether certain variations follow a probabilistic model distribution is a very challenging problem. Agriculture is

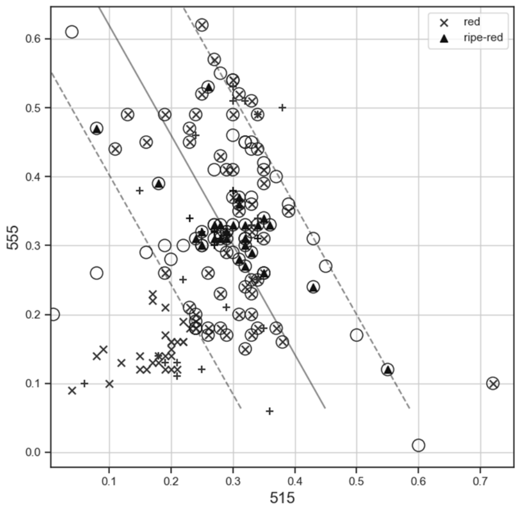

Figure 4. Show Spectral Distribution of Selected btween Light Wavelength in Visible light regin as Pairplot.

Note: The units for both the vertical and horizontal axes are [nm].

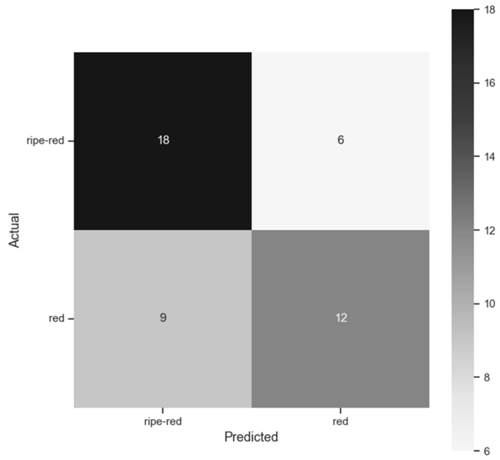

Figure 5. Show Confusion Matrix between growth red Tomato and ripe-red Tomat with Spectral Distribution for model suitability.

Note: Each axis represents the frequency, and the numbers in the boxes represent the degree of fit.

governed by the natural environment, and in controlled environments, such as facility cultivation, environmental control is influenced by external factors and disturbances that affect surrounding changes. Additionally, human errors during sampling can be expected. It cannot be assumed that all sampling groups that distinguish similar items are correct, as subjective judgment by the measurement operator may be involved. Therefore, in this case, preprocessing is needed to remove outliers based on a statistical criterion for identification within the point cloud. In this instance, we used the interquartile range (IQR) from the box plot as the upper and lower limits for determining and removing outliers.

IV. EXPERIMENT.

A. *Experimental Field*

We conducted an experimental farm for tomato cultivation incorporating the UECS environment as the experimental field. Each year, new crop experiments are established by re-creating the beds around the summer Obon period. The beds utilize a coconut coir substrate for solution cultivation, allowing the roots to absorb moisture and liquid fertilizer. Since pesticide is regularly sprayed once a week, we used the prototype to conduct light multispectral measurements on the fruits and leaves just before pesticide application.

B. *Measurement Timing*

In this experiment, we measured the light spectrum according to the color of the fruits after the tomato seedlings had grown and the height exceeded 1 meter, approximately five weeks after planting, when the lower and middle sections of the plant bore fruits that changed color from green (immature) to red (orange yellow) (see Figures 1 and 2). The number of fruits measured was approximately 70 for each color, and additional fruits are expected to develop in the upper section, with green fruits turning orange-yellow and orange-yellow fruits turning red.

C. *Measurement Results and Discussion*

According to the light multispectral distribution measured for the green and red fruits, the normalized reference wavelengths showed a maximum of 445 nm and a minimum of 630 nm, indicating that the differences in other wavelengths reflect the variations in fruit growth. The prominent trends observed in the growth process from green to red fruits are found at wavelengths of 415 nm and 515, 555, and 680 nm. The wavelength of 415 nm corresponds to the blue region of visible light, and as the fruit grows and ripens, the light intensity increases. In contrast, the wavelengths corresponding to the yellow-green region, 515 and 555 nm, both show a decrease in light intensity, and the wavelength corresponding to the red region, 680 nm, also decreases in intensity. This suggests that measuring the light multi-spectrum of the target crops can be linked to the color judgments made by nearby agricultural workers. Traditionally, the near-infrared range (900 nm and above) has been used along with the red wavelength (680 nm) to calculate the Normalized Difference Vegetation Index (NDVI), which indicates vegetation vitality. This discussion focuses on grouping visible light wavelengths and exploring how the distribution of light intensity changes during the growth process of vegetation, which leaves ample room for further discussion.

This complexity arises because discussing light wavelength groups in relation to time and vegetation growth is challenging in a simple planar matrix or three-dimensional space.

D. Classification Using Machine Learning

In this experiment, we use tomato growth in facility cultivation as the sample. We select two wavelength components as features from among eight light spectra in the visible light range during the ripe and harvest periods (Figure.3). For the pair of wavelengths (515-555 nm) selected as classifiable from three distributions, including the blue period, we performed classification using a Support Vector Machine (SVM). The parameters used for the classification model were the cost parameter C (=100) and training data T (=0.2), allowing for data to be included within the margin and classification boundary. As part of the preprocessing for learning, we restricted outlier values that should be judged as anomalies among the factors of measurement variation to those within the interquartile range of the box plot, excluding those outside this range. This enabled classification and exploration of classification boundaries, as compared to using the dataset directly without preprocessing (Figure.4).

Moreover, as an evaluation metric for the performance of the classification model, the fitness assessment from the confusion matrix indicated that while there were a certain number of false negatives and false positives in such cases, the precision could be maintained above 50% (Figure.5). This suggests that when the dataset used for training includes outliers, it becomes impossible to classify them on a two-dimensional plane coordinate if they are contained within other datasets, leading to limitations in the two-dimensional space. To address the distribution of these two-class point clouds, an effective method involves extending interpretations to a certain rule, allowing for a transformation of the two-dimensional plane into a three-dimensional linear space, thus achieving dimensional separation at the boundary surface. However, it is generally acknowledged that there is no universally applicable formula for generalization and that the computational time for learning increases significantly with dimensionality increase, although this discussion is omitted in this paper. It is important to note that raw data used in actual operations almost invariably contains missing values and anomalies, which can hinder the learning of classification models. Therefore, we believe it is crucial to establish a situation where preprocessing is necessary to facilitate learning while maintaining a lower dimensionality.

V. CONCLUSION.

This study presents the results of experiments conducted using a prototype measuring instrument to quantify the growth degree of crops during the growth process through visible light multispectral (light wavelength) analysis. The enhancement of crop growth and sweetness quality is influenced by many complex conditions, including not only photosynthesis from sunlight but also factors such as the duration of sunlight and the timing of water application. Furthermore, we demonstrated that preprocessing of data is essential when applying machine learning using Support Vector Machines (SVM) to classify and optimize the classification of growth degrees, particularly when dealing with raw data that may contain outliers. In the future, it is expected that the quantified light spectrum distribution information can be utilized as additional information for sorting machines and automated harvesting robots, leading to rapid and accurate classification.

REFERENCES

[1] Murakami :"Morphological Control of Plant Growth in Protected Cultivation by Changing Optical Radiant Environment", Journal of the Illuminating Engineering Society, "Plant cultivation techniques using artificial light", Vol.79, No.4 p.149-154 (1994)

[2] Takashi Kameoka, Ito, Shin. Kameoka, Hashimoto : "Development of lettuce freshness measurement method using a multi-spectral sensing", The 31th Annual Conference of the Japanese Society for Artificial Intelligence, 2E3-OS-36a-3, (2007).

[3] Miura, Watanabe, Asai : "Sensing of laser Normalized Difference Vegetation Index(NDVI) for vegetation lidar", laser-sensing Society, (2019)

[4] Tao Li, Qichang Yang : Advantages of diffuse light for horticultural production and perspectives for further research, frontiers in Plant Science, (2015).

[5] "Detection of Ripe and Raw Tomatoes using Internet of Things", C.N. Vanitha, 7th International Conference on Computing Methodologies and Commnications (ICCMC-2023)

[6] "Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009." U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2