# An IoT Stereo Image Sensor System for Agricultural Application

Bruno M. Moreno[1,2], Paulo E. Cruvinel[1,2]

[1] Embrapa Instrumentation (CNPDIA), P.O. Box 741, 13560-970, São Carlos, SP, Brazil
[2] Postgraduate Program in Computer Science, Federal University of São Carlos, São Carlos, SP, Brazil
Emails: bruno.moreno@estudante.ufscar.br; paulo.cruvinel@embrapa.br

***Abstract*—Sensors and Internet of Things systems have become quite important to support decision-making in agriculture. In such a context, smart farming has emerged as a new opportunity for food production based on a sustainable development concept, since the rational use of agricultural inputs is now a reality. One of these opportunities is the application of precision agriculture for weed control. This paper presents the characterization of an embedded stereo system using camera sensors, Internet of Things principles for computational intelligence tasks. For validation, it has been used the Modular Transfer Function concept, that is, taking into account not only the calibration of the sensors, but also of the 3D system, memory use and energy consumption for a long term operation. Furthermore, the results clarify details related to the implementation and construction of such a 3D system, which in fact aims to control invasive plants in agricultural crops.**

***Keywords*—camera sensor; stereo vision; embedded platform; IoT sensors; agricultural industry.**

## I. Introduction

Agriculture is a very important source of food, feed, fiber and even fuel. Despite this, agriculture currently faces the challenge of increasing its production in response to the demand of continued population growth, taking precautions against the various adversities caused by the climate and minimizing the impact of man on nature.

Recently, a previous study regarding an Internet of Things (IoT) system for agricultural application was presented at the Eighth International Conference on Advances in Sensors, Actuators, Metering and Sensing (ALLSENSORS 2023) [1].

IoT devices have been used in agriculture, mainly in tasks that aim to reduce waste of resources. As examples of applications, IoT can help with the storage of agricultural products, smart irrigation, soil monitoring, nutrient management, precision agriculture, intelligent livestock management and crop monitoring. In irrigation, devices can automate the process intelligently, collecting data from the soil with temperature and humidity sensors, and using the collected and historical information to train a model to decide the best time to activate irrigators. Other information can be collected from the soil by sensors, such as pH and nutrient content, allowing the choice of the best plant breed for certain soil parameters. This information can be controlled and monitored remotely via web or mobile applications. The sensors can also track the farm and with the data collected, farmers can plan their farming activities such as seed selection, sowing, amount of fertilizer used, harvest date and expected yield amount [2].

One of the approaches aimed at increasing productivity in the field is the reduction of losses due to factors exogenous to crops, such as competition resulting from the presence of invasive plants. The presence of weeds in the cultivation area can decrease crop yield by more than 50% just by competing with the moisture present in the soil, causing more damage than invasive animals, diseases and other pests [3]. Therefore, weed control is essential so that the nutrients present in the soil, the development space and the reception of sunlight remain exclusively for the plant of interest [4].

Moreno and Cruvinel presented previous studies related to a stereo camera's system [5], and the development of a software based on semantic computing concepts for the segmentation of weed plants [6]. Even though there are systems that perform plant phenotyping [7], none have combined the information generated by stereo images, as the system developed can also provide the height of the plants as data to assist in the task of deciding the correct quantity of product in the region. Of the works that use more than one camera to obtain images, there is a greater focus on 3D reconstruction of plants, which allows generating a point cloud representation of the plant from a depth map [8].

Although the use of pesticides has already been established to deal with this problem, technological applications aimed at the rational use of inputs are desired. Among such technologies, Computer Vision stands out, which works in two stages: image acquisition and image processing. The acquisition is made exclusively from camera sensors, capturing the environment and patterns present in digital images. Such sensors can then capture the visible or thermal spectrum, and be coupled to vehicles, devices, robots, drones and even satellites. On the other hand, affordable single-board computers have made onboard image processing possible [9].

Image processing, a task of computacional intelligence, can be summarized in five steps. In the first, the raw data are preprocessed, removing noise and selecting only the object of interest. In another step, pattern features are extracted, whereas in the case of plant images, such parameters are related to color, shape and texture. In the third stage, the features go through a selection process, decreasing the dimensionality of the data. Afterwards, the data are classified, grouping them based on their similarities. Finally, in the decision-making stage, new input data can be classified from the already trained model, thus identifying which group it belongs to [10][11].

To ensure that the input data are of good quality, validating and using good camera sensors have become extremely impor-

tant. Allied to this, other points of consideration in the application of such techniques in agriculture are the management of the volume of data generated, the data analysis techniques that need to deliver interpretable and understandable results due to the interdisciplinarity of workers in the area, and the mobile systems that need to be able of handle scarce resources such as limited battery life, low computational power and limited bandwidths for data transfer [12].

As examples of the use of camera sensors in the field, there are applications coupled to vehicles to operate during preplanting and analyze the height and density of vegetation from the images [13] and identify the location of invasive plants for manual control via weeding machine [14]. Plant images can also be acquired to create a database on an external server for further processing, for training a future classifier [15].

This paper is structured as follows. Section II presents the materials and methods used, including the IoT system description, camera sensor specifications, stereo vision basics, and embedded board specifications. Section III presents the results of the validation of the sensor and of the stereo system, the power supply and memory limitations and the final prototype, with the final conclusions in Section IV.

## II. Materials and Methods

The developed system aims to capture stereoscopic images in a real environment of plantations, so that the presence and concentration of weeds present in the region of interest can be identified from an embedded algorithm. The capture of stereo images requires two camera sensors, generating two images of the same area that will be the input of the system. The images are then processed and grouped into classes, and the data will be prepared for sending to a module external to the system, which will be responsible for spraying the site.

### A. High-level IoT architecture

Embedded systems have a potential in agricultural use due to their mobility, low cost and computational power, allowing the performance of complex tasks in a more practical way. Raspberry Pi (RPi) is being used in several applications and it is the leading candidate for hardware implementation due to its powerful processor, rich I/O interface and compatibility that allows most projects to run on it [16]. Its wireless communication also makes the RPi capable of working with IoT projects, allowing objects to be sensed or controlled remotely across existing network infrastructure and reducing human intervention [17].

IoT systems in agriculture are separated into three modules: farm side, server side and client side. The farm side usually consists of detecting local agricultural parameters, identifying the location and sensor data, transferring crop fields data for decision-making, decision support and early risk analysis based on recent data, and action and control based on the monitoring of the crop [18]. As can be seen from the block diagram in Fig. 1, the farm side is represented by the developed IoT Stereo System that can gather image data in the field, pre-process, segment, create feature extraction and depth information vector, classify and interpret the collected data, while being controlled and monitored via Bluetooth serial communication by a mobile app.

On the server side, the network layer is responsible for reliable transformation to the application layer. A Wireless Personal Area Networks (WPAN) network can be mounted on a single board computer, with its own unified control and monitoring console for various wireless networks. Data transport and storage become essential, with data that can be saved on an external server or in the cloud, and then transferred to other devices, including the equipment responsible for product spraying on the plantation. The last module, the client side or application layer, collects and processes information, providing an environment where users can monitor data processed by the system via a web browser, anywhere and anytime. In Fig. 1, the server side is represented by the Bluetooth and Wi-Fi communication of the system to the farm server and by the server management of local and remote network, while the client side is represented by the mobiles devices and by the cloud environment.

Communication between all devices can be carried out via the Bluetooth protocol, which supports up to 7 devices connected simultaneously. The Bluetooth 4.2 connection can reach the transfer limit of 1 Mbps and the signal can reach 10 m away from the board indoors and 50 m outdoors. One of the protocols used is radio frequency communication (RFCOMM). The RFCOMM protocol is an important layer that provides a serial interface to the Bluetooth transport layer, emulating an RS-232 interconnect cable. RFCOMM is based on the ETSI 07.10 standard, which allows the emulation and multiplexing of multiple serial ports on a single transport [19]. The OBEX protocol (OBject EXchange) is also used for file transfer, which is a software implementation of the File Transfer Protocol (FTP) network protocol, which runs on top of RFCOMM.

To ensure system security, it only connects to trusted equipment and specific ports. An RPi is then defined as master, responsible for receiving commands sent by an application on an Android cell phone and using this command to carry out its actions and inform the other board what it should also do. The other RPi is defined as a slave, receiving commands from the master and obeying them.

The pseudocode of the algorithm developed for the system communication between all components is described below, where $addr\_master$, $addr\_slave$ and $addr\_mob$ are the MAC addresses of the master RPi, slave RPi and mobile controller, respectively, and $prt\_1$ and $prt\_2$ are the ports enabled for serial communication:

---

**function** MASTER COMMUNICATION($addr\_slave$, $addr\_mob$, $prt\_1$, $prt\_2$)
**begin function**
$s1$ = create_socket_bluetooth(RFCOMM)
$s2$ = create_socket_bluetooth(OBEXFTP)

```
    connect(s1,(addr_slave, prt_1))
    bind(s2,(addr_mob, prt_2))
    while mobile connection is not interrupted do
        comd = get_data(mobile)
        send(comd,slave)
    end while
    return comd
end function
function SLAVE COMMUNICATION(addr_master, prt1)
begin function
    s = create_socket_bluetooth(RFCOMM)
    bind(s,(addr_master, prt_1))
    accept_conection(s)
    while master connection is not interrupted do
        comd = get_data(master)
    end while
    return comd
end function
```

The system is built in such a way that it can be operated in the field without the need for an Internet or 5G connection, which allows the data collection stage to work in more isolated locations, requiring only the existence of a local network. The processed data can be used by other devices connected to the local network, but can also be transmitted to external servers later when the connection to the World Wide Web is available. In this way, specific commands can be sent remotely by the user to the system, which will perform procedures such as image capture and data transfer.

### *B. Embedded System and Camera Sensor Specifications*

RPi is a series of mini-embedded computers developed in the United Kingdom by the Raspberry Pi Foundation in association with Broadcom. The model used was the RPi 3 B+, where its specifications can be seen in Table I. It is important to note that the board must be powered with a nominal voltage of 5 V capable of delivering 2.5 A of current, with operating temperature between 0 °C and 50 °C.

The internal memory is defined from a micro Secure Digital (SD) card, where the kernel of the operating system is also present, being recommended the use of at least 8 GB of memory. The RPi 3 B+, unlike previous family models, enables BCM43438 wireless Local Area Network (LAN) and Bluetooth Low Energy (BLE) communication, allowing wireless data exchange.

The RPi has its own camera sensor alternatives, including the Pi Camera v1, with specs shown in Table II. Among the most important parameters, stand out the fixed focal length of 3.60 mm, the maximum sensor resolution of 2592 x 1944 pixels, and the camera opening angle of 53.50º horizontally and 41.41º vertically.

### *C. Modular Transfer Function as Camera Sensor Validation*

Lens and camera designers face challenges in developing systems with high image quality. The problem of greatest concern is how to optimize lens parameters such as curvatures and thicknesses to obtain high image resolution. A set of optimizations were proposed to improve the aberrations of lens systems, using as a metric the Modular Transfer Function (MTF), which is the amplitude term of the Optical Transfer Function (OTF) which is similar to the transfer function of the linear system [20]–[22].

To evaluate the quality of the images acquired by the cameras, the MTF is used from each one of them and from the set, expressing how well an optical system preserves the contrast of spatial frequencies of the object in the image and is a well-established performance method [23].

A simple method to obtain the transfer function is to generate the system response when the input is a pure impulse signal, therefore obtaining the impulse response of the function. Using the same procedure, a point source is considered as the impulse signal to help estimate the image response in a lens system.

The point source image shown on the image plane is called the Point Spread Function (PSF), which is the inverse Fourier transform of the OTF. The projection of the PSF in 1D is called the Line Spread Fuction (LSF), measurement preferable because it can be obtained simply and equally valid for cases where there are no distortions between the axes.

Then, the camera sensor can be defined taking into account the calculation of the LSF of the camera lens and the MTF, which represents the magnitude response of the optical system to sinusoids of different spatial frequencies, that is, recovered by the Fourier transform of the LSF. Several key aspects of optical instrumentation relate to the implementation of a linear source for a given optical system, the impact of finite source size on measurement, and the choice of optical elements for imaging the response of specific patterns and their relationship to the lens used in the camera sensor.

Taking a linear source, the solution to measure the MTF is in 1D orthogonal to the direction of the line. This can be proven considering a given source $S(x, y) = \delta(x)C$ and a lens with a diameter equal to $a$, obtaining the objective response $R(k_x, k_y)$, described in Equation (1).

$$R(k_x, k_y) = \int \int_{-a/2}^{a/2} \delta(x) C e^{j(k_x x + k_y y)} dx dy \qquad (1)$$

Thus, the spatial frequencies associated with the spatial coordinator $(x, y)$ can be expressed as the square of the Fourier transform of the product of the source and lens aperture $R^2(k_x, k_y)$, with $(k_x, k_y)$. Therefore, looking for the solution of (1) and solving the integral by parts, it is possible to arrive at:

$$R^2(k_x, k_y) \alpha \frac{sin^2(ak_y)}{(ak_y)^2} \qquad (2)$$

Equation (2) corresponds to the LSF. The Fourier Transform of the LSF then gives the 1D MTF in the yy direction. Considering that the lens has circular symmetry, using this function it is now possible to characterize the entire lens.

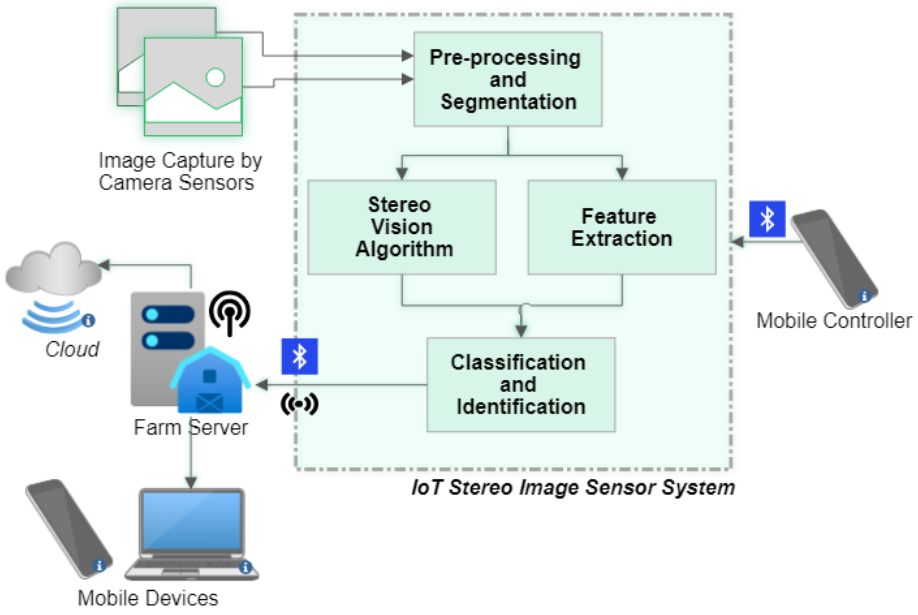



Figure 1. High-level system architecture diagram.

TABLE I
RASPBERRY PI 3 MODEL B+ CHARACTERISTICS

| | | | |
|---|---|---|---|
| Processor | BCM2837B0 Cortex-A53 (ARMv8) 64-bit | | |
| Clock | 1.4 GHz | GPIO | 40 pins |
| Memory | 1 GB SDRAM | Gigabit Ethernet | 1 connector |
| USB Port | 4 USB 2.0 | HDMI | 1 connector |
| Camera serial interface (CSI) | | Display serial interface (DSI) | |
| Wireless (dual band) | | Bluetooth 4.2/BLE | |
| 3.5mm 4 Jack output | | Micro SD card slot | |
| Support Power-over-Ethernet | | Input DC 5V/2.5A | |

TABLE II
PI CAMERA CHARACTERISTICS

| | |
|---|---|
| Size | 25 x 24 x 9 mm |
| Resolution | 5 MP |
| Video modules | 1080p30, 720p60, 640x480p60/90 |
| Sensor | OmniVision OV5647 |
| Sensor resolution | 2592 x 1944 pixels |
| Sensor image area | 3.76 x 2.74 mm |
| Pixel size | 1.4 $\mu$m x 1.4 $\mu$m |
| Optical size | 1/4" |
| Full-frame SLR equivalent | 35 mm |
| S/N Ratio | 36 dB |
| Dynamic range | 67 dB @ 8 times gain |
| Fixed focus | 1 m - $\infty$ |
| Focal length | 3.60 ± 0.01 mm |
| Horizontal field of view (HFOV) | 53.50° ± 0.13° |
| Vertical field of view (VFOV) | 41.41° ± 0.11° |
| Focal ratio (F-stop) | 2.9 |

A popular way of estimating the MTF curve for spatial frequency is called the inclined knife-edge method, in which the curve is obtained from a region of the image where there is a transition from a very dark tone to a very light tone [24]. An Edge Spread Function (ESF) is calculated from the recorded knife edge, giving the unidirectional response of the imaging system to an edge object, replacing the PSF. The LSF can then be obtained in the same way from the derivative of the ESF and finally the MTF is calculated from the Fourier Transform.

In stereo systems, the system MTF is generally summarized as a set of curves for each sensor used, or just the curve of the lowest quality sensor [25]. In this research, the response of all sensors is considered, performing the convolution of the sensors' responses, based on the multiplication of the MTFs in the frequency domain, as illustrated in Equation (3).

$$MTF_{sistema} = \mathcal{F}(LSF_1 * LSF_2) = MTF_1 \times MTF_2 \quad (3)$$

To qualify a sensor, three points of the MTF are usually analyzed: the frequency at which it drops by 50% (at which the image contrast is degraded by half), the frequency at which it drops by 10% (at which the image contrast is image is degraded by 90%) and the MTF value at the Nyquist frequency, which should preferably be greater than 0 [26]. Considering these aspects, the MTF becomes fundamental in analyzing image contrast, so that the impact of spatial resolution and lighting variations can be analyzed. If contrast is compromised, texture and edge details of plants may be damaged to the point of making it impossible to extract features correctly.

Figure 2 shows an example of the typical images where the weed identification task can be performed and the expected size of the plants present. For such situations, the MTF itself can be used in image enhancement processes, based on the deconvolution of the signal based on a Wiener filter [27]. The characterization of the MTF is then useful to define the spatial response of the vision system, considering its detection capacity from a minimum dimension in pixels of the object of interest.

The pseudocode of the system MTF calculation algorithm developed, with the left image $I_L$, right image $I_R$ and

number of samples $n$ as inputs, can be described as:

**function** SYSTEM MTF CALCULATION($I_L$,$I_R$,$n$)
**begin function**
  **for** each image $I_L$ and $I_R$ **do**
    Form $n$ *subimages* from regions where there is an inclined knife edge
    **for** each *subimages* $n$ **do**
      $ESF(n)$ = read_value_pixels(centered horizontal line)
    **end for**
    $ESF$ = average($ESF(n)$)
    $ESF$ = normalize($ESF$)
    $LSF$ = derivative($ESF$)
    $MTF$ = Fourier_transform($LSF$) ▷ from $I_L$ obtain $MTF_L$, and from $I_R$ obtain $MTF_R$
  **end for**
  $MTF\_system$ = $MTF_R \times MTF_L$
**return** $MTF_R$, $MTF_L$ and $MTF\_system$
**end function**

With the three MTFs, it is then possible to validate the sensors individually and together in the system.

### D. Stereo Vision Principles

Stereo vision systems are usually based on the use of two cameras with the aim of simulating the human vision system and obtaining depth of objects, with the camera plane as a reference. The depth is acquired through the comparison of the object's position between each captured image [28]. The simplest way of comparing both images is guaranteed when the cameras are coplanar and aligned, as shown in Fig. 3. The variables defined by the camera system are the baseline $b$ and the focal distance $f$. The $P(X, Y, Z)$ represents a point that would be recorded by the two cameras and $u_L = (X_L, Y_L)$ and $u_R = (X_R, Y_R)$ are the projections of this point in each image. From the concepts of geometry and similarity of triangles, it is possible to obtain:

$$Z = \frac{bf}{X_L - X_R} = \frac{bf}{d} \tag{4}$$

The $d$ variable is called disparity. Thus, with two images as inputs in a calibrated and synchronized stereo architecture, depth information is obtained by finding the corresponding pixels in both images ($u_L$ and $u_R$) by a matching algorithm and subtracting their X-axis coordinates. By performing this operation for all paired pixels in the image, the disparity map is obtained, which contains all the depth information in the image.

It is also important to note the distortion that variations in the disparity map can cause in the depth estimation, i.e., verify the measurement obtained accuracy. So, for a variation in depth, it is possible to find:

$$\Delta Z = Z - \frac{bf}{d + \Delta d} = \frac{Z^2 \Delta d}{bf + Z\Delta d} \approx \frac{Z^2 \Delta d}{bf} \tag{5}$$

Therefore, when designing a stereo system, attention must be paid to a baseline value at which objects at the distance of interest can be correctly differentiated while the measurement depth distortion must be small. Another important factor when designing such a system is the calibration of and between cameras.

For camera calibration, the set of internal parameters is considered to validate the method. Every camera can be described based on intrinsic and extrinsic parameters, which contribute to how the image is formed from the scene in the real world.

The intrinsic parameters are those related to internal biases, due to the sensor and its shape, the lens and its distortions and other characteristics involved in the manufacture of the camera, while the extrinsic parameters refer to the position of the camera in space in relation to the world. The extrinsic parameters can be simplified by a rotation matrix $\mathbf{R_m}$ and a translation matrix $\mathbf{T_m}$ [29].

The focal length $f$ is an intrinsic parameter, as it is the distance between the center of the camera and the image plane, i.e., from the lens to the censor. Many cameras use a Charge-Coupled Device (CCD) sensor, a semiconductor sensor formed by an integrated circuit that contains a matrix of coupled capacitors, capable of generating electrical stimuli from the light received. As the pixel on a sensor of this type may not be perfectly square, there is the possibility of a small distortion in the number of pixels per unit length. In this way, the focal length of the camera lenses will be different in each direction, resulting in the variables $f_u$ and $f_v$, with the aspect ratio being defined by $f_v/f_u$.

Another camera parameter is the optical center, defined by the coordinates $(u_0, v_0)$, which represents a translation factor of the image origin in relation to the center of the sensor, such that the image origin is correctly on the upper edge left of her. There is also the skew coefficient ($\tau$) that corrects the image in cases where the CCD sensor does not have a perpendicular orientation between the length and width axes. As this situation is rare for most sensors, it is common to assume that $\tau = 0$.

Finally, due to the curved nature of lenses, the last intrinsic parameters to be considered when modeling a camera are the distortion coefficients [30]. The tangential distortion coefficients are defined by two variables, $k_{p1}$ and $k_{p2}$, while the second, fourth and sixth order radial distortion coefficients are respectively represented by $k_{q1}$, $k_{q2}$ and $k_{q3}$.

Therefore, the process of capturing a digital image by a sensor can be described in a simplified way using Equation (6), based on the projection of space onto the sensor, where $u_d$ and $v_d$ represent the coordinates of a point in the image without distortion correction, $s$ the scale or resolution factor and $X_w$, $Y_w$ and $Z_w$ the coordinates of a point in the world.

Figure 2. Typical images of plants in crops for phenotyping task.

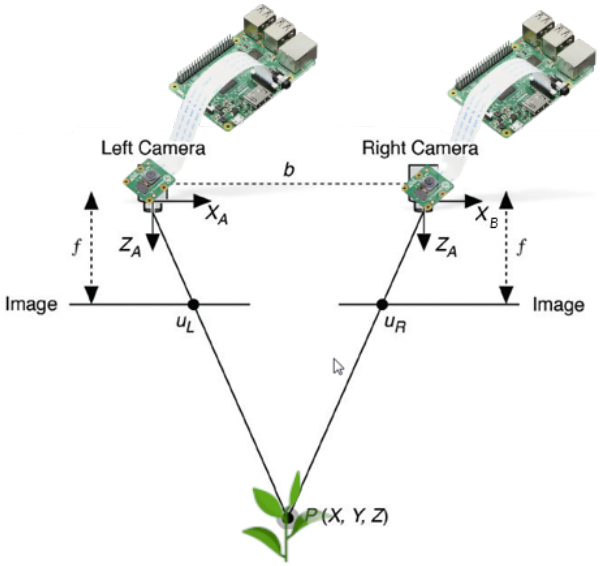



Figure 3. Stereo vision model.

$$s\begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \tau & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R_m} & \mathbf{T_m} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (6)$$

To find the undistorted coordinates $(u, v)$ of the image, correcting the projection, the system of equations (7) must be solved.

$$\begin{cases} x_d = \dfrac{u_d - u_0}{f_u} \\ y_d = \dfrac{v_d - v_0}{f_v} \\ r^2 = x_n^2 + y_n^2 \\ x_k = x_n(1 + k_{q1}r^2 + k_{q2}r^4 + k_{q3}r^6) \\ x_d = x_k + 2k_{p1}x_n y_n + k_{p2}(r^2 + 2x_n^2) \\ y_k = y_n(1 + k_{q1}r^2 + k_{q2}r^4 + k_{q3}r^6) \\ y_d = y_k + 2k_{p2}x_n y_n + k_{p1}(r^2 + 2y_n^2) \\ u = f_u x_n + u_0 \\ v = f_v y_n + v_0 \end{cases} \quad (7)$$

In addition, when characterizing the intrinsic parameters of any camera, the information can be summarized from two matrices, the camera matrix $\mathbf{M_{cam}}$ and the distortion coefficient matrix $\mathbf{K_{cam}}$, as can be seen in Equations (8) and (9).

$$\mathbf{M_{cam}} = \begin{bmatrix} f_u & \tau & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$\mathbf{K_{cam}} = \begin{bmatrix} kq_1 & kq_2 & kp_1 & kp_2 & kq_3 \end{bmatrix} \quad (9)$$

The process of obtaining such parameters is called camera calibration. Calibration methods depend on the model used to approximate actual camera behavior. The most used models are the linear models of Hall and Faugeras–Toscani, developed respectively in 1982 and 1986, and the non-linear models of Tsai and Weng, implemented in 1987 and 1992, which generally present fewer errors [31].

From Equation (6), the projection matrix between the real world and the image universe will have dimension equal to 3x4, which results in 11 parameters that must be obtained. Commonly, to calibrate stereo systems and obtain matrix

values, images of chessboards with known dimensions (number of squares and size of their side in the real world) are used, in which the calibration points are the internal vertices of the squares on the board. As each point corresponds to two equations (one in the x coordinate and the other in y), five and a half points are needed to calibrate the system, but experiments have shown that 5 times more points than necessary gave better results [32]. A greater number of images for calibration also reduces the total location error in mm, with 13 images in which at least 30% of them were composed of the chessboard, in random orientations, already shows good results. Calibration methods other than the board can also be used, such as calibration using a laser [33] or using spherical objects [34].

The entire stereo vision system must also be calibrated, where are obtained the rotation factor $\mathbf{R_{stereo}}$ and the translation factor $\mathbf{T_{stereo}}$ between the left and right image. For this calculation, the previously calculated camera parameters and the simultaneously captured chessboard images are used. The process of correcting the orientation of stereo images is called rectification. Note that, unlike the camera matrix and distortion coefficients which depend only on the camera, the $\mathbf{R_{stereo}}$ and $\mathbf{T_{stereo}}$ matrices must be recalculated if any stereo system settings change such as, for example, the baseline distance.

## III. Results and Discussions

Experimental results were focused on the instrumentation's characterization, i.e., including both the sensors and hardware associated with signal and image processing. So far, the images for such a characterization were collected at laboratory level only. The system is based on eight elements, as follows: 12 V battery; 12 Vdc to 220 Vac voltage inverter; Light Emitting Diode (LED) lamp; 110-220 Vac to 5 Vdc rectifier; two RPis and two Camera Pi, as the schematic presented in Fig. 4. All components are fixed on a metallic structure, with adjustable distance between cameras, angle of inclination (0º, 90º, 180º, 270º) and height of the cameras in relation to the ground (10 to 100 cm). The constructed system can be seen in Fig. 5.

The system is controlled by an Android App via Bluetooth serial communication, where commands can be sent: synchronous image capture on the two RPis, send the images to the cell phone to check the quality of the capture, check the amount of images saved on memory, and board reboot or shutdown command. The RPis also communicate with each other via Bluetooth protocol, that supports up to 7 accessory devices, and uses RFCOMM Bluetooth protocol in data transfer with the cell phone. To ensure system security, it connects only to trusted equipment from their MAC address on specific designated ports.

The elements that most impact the cost of the system are those related to the power supply, sensors and the embedded board. The advantage of the RPi is that it is cheaper when compared to other boards such as the PC/104, although a more detailed analysis should take into account local and freight costs and component availability.

### A. Energy consumption management

A RPi can have power consumption of up to 12.5 W, but in laboratory tests the usual value during the application of the image capture software was only 3 W. As the system was designed with an inverter, the power consumed by this equipment must also be considered for system evaluation and possible improvements. In this case, the inverter in question presented a spent power of around 8.4 W, significantly higher than the sum of the RPis. To deal with such power, a battery of 12 V and 60 Ah was chosen.

To measure the energy expenditure of the system, current and power were calculated in different situations, according to Table III, with battery voltage fixed at 12.0 V. To evaluate the battery capacity, a test was carried out in the most extreme situation, with the system in continuous operation with the 18 W LED lamp on, which resulted in the maintenance of operation for approximately 15 hours. When the battery was discharged to 11.7 V, the inverter stopped as a safety precaution. It was observed that, in this operating mode, the peak current at system startup was close to 3.2 A, while with the same configuration but with the less potent LED lamp the peak was 2.0 A.

### B. Memory management

For each RPi a 32 GB SD memory card was selected. After the initial settings, the necessary programs installed and the capture algorithm developed, about 23.1 GB of memory was free for general use. To ensure that the program can handle the amount of data written and stored, it is good to know how long the embedded system takes to save files. In testing, it was found that the SD card sequential memory write rate is 14833 KB/s or 14.8 MB/s.

Such information is important to define the resolution in which the images will be captured, as they define the size of the files saved in memory. Following the dimension of the camera sensor, it is preferable to define the resolution of the captured images to take advantage of the entire sensor size, that is, in which the 4:3 ratio is preserved. The maximum file size can be calculated by multiplying the resolution by the pixel depth, but since the Pi Camera doesn't have the option to format a RAW image file, the images are compressed, resulting in smaller files. So, it was tested in five resolutions, 640 x 480, 800 x 600, 1024 x 768, 1280 x 960 and the maximum 2592 x 1944. Early test results can be seen on Fig. 6, where five images in each resolution were taken and saved in the PNG format.

Considering the future application in image processing, in which the computational cost of operations tends to grow exponentially according to the number of pixels present, and the available SD memory, the resolution of 1280 x 960 was then chosen. With this resolution, at least 6,000 images can be saved in memory, although it is possible to store them later in the cloud, from the system's communication with an external network, freeing up space on the board.

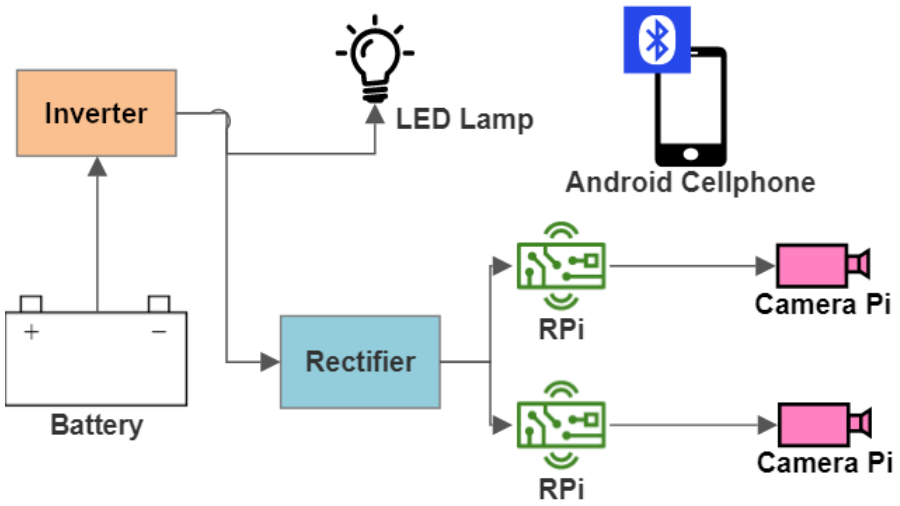
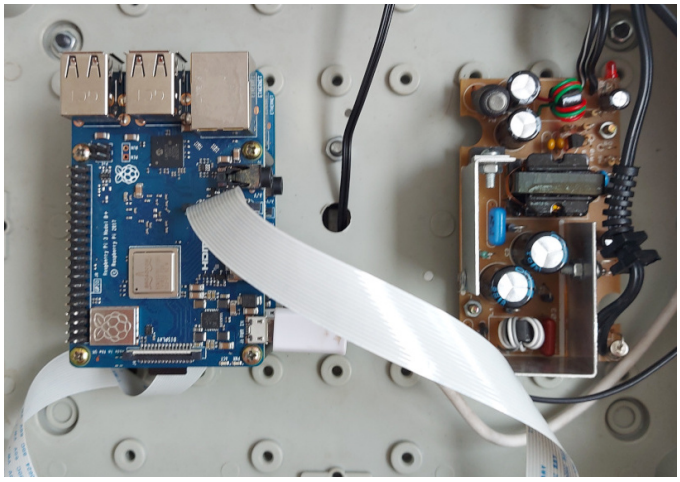



Figure 4. Diagram of the connection between the components.

(a) Details of the camera, stereo rig and lamp.



(b) Interior of protective case, with RPis and rectifier.

Figure 5. Developed system.

TABLE III
SYSTEM POWER AT DIFFERENT SETTINGS

| Mode of operation | Current (A) | Power (W) |
|---|---|---|
| Standard | 1.3 | 15.6 |
| With active camera sensors | 1.4 | 16.8 |
| With active camera sensors and 4.5 W LED lamp | 1.9 | 22.8 |
| With active camera sensors and 18 W LED lamp | 3.0 | 36 |

It should be noted that for future applications, if the maximum resolution is used, the memory writing time must be taken into account as a limiting factor.

### C. Camera sensor validation

The first step in calculating the stereo MTF was to capture an image of the chessboard with both cameras at the same time, as can be seen in Fig. 7. For each image, five random regions were selected where there are knife edges recorded in the same location for both cameras. The ESF and LSF for a sample of the left and right camera can be exemplified in Fig. 8. The normalized MTF was calculated for each point and averaged between them.

The MTF for the stereo system was then calculated from the convolution in the frequency domain of both partial MTFs, i.e., each one obtained for the cameras used in the developed stereoscopic image system (Fig. 9). For the left camera, 50% of contrast reduction was observed for the normalized special frequency equal to 0.327 cycles/pixel; and 90% of contrast reduction at the 0.551 cycles/pixel. For the right camera, the values of reduction in such frequencies were respectively 0.286 and 0.673 cycles/pixel. Besides, for the entire system, the 50% of contrast could be found in the 0.224 cycles/pixel and the 10% in the 0.367 cycles/pixel. Therefore, the MTF value at the Nyquist frequency was equal to 14.31% for the left camera, 8.97% for the right, and 1.28% for the entire stereoscopic

system. As the MTF value was greater than 5% (contrast reduction that still allows the recovery of the edges of the objects in low noisy images), as well as greater than 0% for both cameras in the system. Such a result qualifies the CCD's sensors, which meet the needs of the developed prototype.

By using the MTF concept, it has become possible to know whether the image will have enough contrast to differentiate the leaves of weed plants when applied in a real agricultural situation. Therefore, considering average values of areas for both weed plants, narrow leaves (monocotyledons) and broad leaves (dicotyledons), the frequencies, in cycles/pixel, could be characterized as 0.053 and 0.100 respectively. Likewise, considering the highest frequency of leaves as a critical point, the MTF presented a value of 97.23% for the left camera, 91.74% for the right and 89.21% for the entire stereoscopic system. Contrast loss values were approximately 10%, which did not interfere with the results, validating the sensor arrangement as suitable to weed family's patterns recognizing.

To evaluate the camera's SNR ratio, only the regions of the converted grayscale image where black blocks were presented, which have a uniform color on the original chessboard, were used, and the mean and standard variation of the signal were evaluated. For the right camera, the calculated value was 19.7 dB, while for the left camera it was 17.9 dB, below the 36 dB specified by the manufacturer.

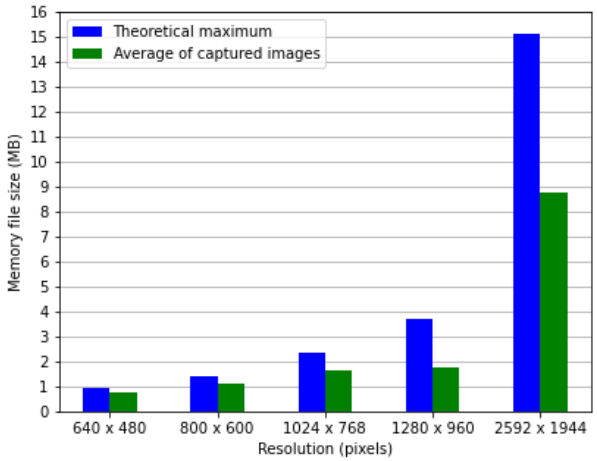



Figure 6. Image file size experimentations.

### D. Stereo vision parameters

The first step in tuning the stereo system is to define the baseline distance that will be used to capture the images. The developed prototype has a minimum possible baseline of 6 cm and a maximum of 24 cm, which makes it capable of simulating human vision, which has this value in the range of 5.4 to 7.4 cm, in addition to allowing the exploration of other scenarios. For this, considering (4) and (5), the expected disparity for an object up to 1 m away from the camera and the expected distortion error at such distance were calculated, for four values of baseline, 6 cm, 12 cm, 18 cm and 24 cm, as can be seen in the Fig. 10, considering the resolution of 1280 x 960.

When setting the baseline distance, it is always preferable to use the lower values to ensure greater interpolation between the two generated images, which allows closer objects to have their distance calculated. For example, according to the graph shown, for b = 24 cm, objects up to 23.8 cm away from the camera would not be present in both images, making it impossible to calculate the disparity, while for b = 6 cm such a situation is only valid for objects less than 5.9 cm away. As for objects of up to 1 m, the distortion error proved to be small for all cases, including for the scenario with the smallest baseline, so it can be defined that the best use of the stereo system occurs for values close to 6 cm.

Thus, for b = 6 cm and height of 1 m (value chosen so that, due to the height of the growing plants, the object under analysis is not too close to the sensors), the calibrated parameters results of the left and right cameras, and of the stereo system, were:

$$\textit{Left camera matrix} = \begin{bmatrix} 736 & 0 & 582 \\ 0 & 735 & 464 \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

$$\textit{Left distortion coefficients} = \begin{bmatrix} 0.0589 \\ -0.169 \\ 0.00139 \\ 0.00198 \\ 0.142 \end{bmatrix}^{\mathsf{T}} \quad (11)$$

$$\textit{Right camera matrix} = \begin{bmatrix} 1480 & 0 & 681 \\ 0 & 1480 & 480 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$\textit{Right distortion coefficients} = \begin{bmatrix} -0.0728 \\ 3.98 \\ 0.00117 \\ 0.00630 \\ -22.6 \end{bmatrix}^{\mathsf{T}} \quad (13)$$

$$\mathbf{R_{stereo}} = \begin{bmatrix} 0.960 & -0.0133 & -0.281 \\ 0.0159 & 1.00 & 0.00721 \\ 0.280 & -0.0114 & 0.960 \end{bmatrix} \quad (14)$$

$$\mathbf{T_{stereo}} = \begin{bmatrix} -0.787 \\ -0.0670 \\ 5.65 \end{bmatrix} \quad (15)$$

Note that if the baseline distance is changed, it is necessary to calibrate the system again, recalculating only the $\mathbf{R_{stereo}}$ and $\mathbf{T_{stereo}}$ matrices, but it is expected that $\mathbf{R_{stereo}}$ will not change significantly, as the mounted structure does not allow the cameras to yaw, pitch or roll.

From these calibration matrices, images can then be correctly rectified, eliminating distortions characteristic of the sensors during image capture and preparing them for use in stereo vision matching algorithms.

(a) Left Camera. (b) Right camera.

Figure 7. Images of a calibration chessboard, captured synchronously and without being processed.

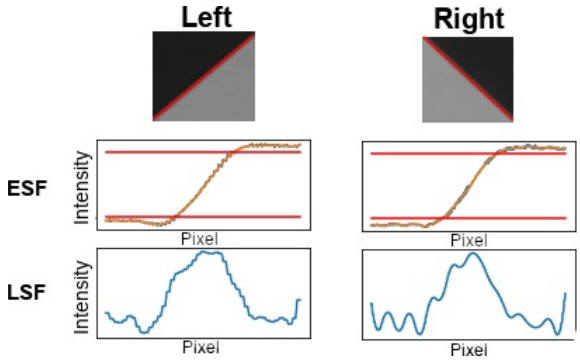



Figure 8. ESF and LSF of a left and right camera sample.

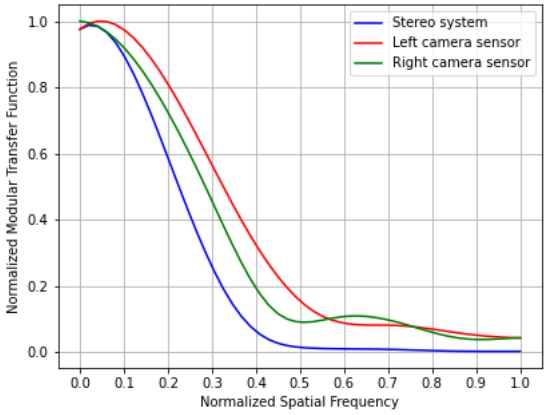



Figure 9. MTF of each camera sensor and combined system.

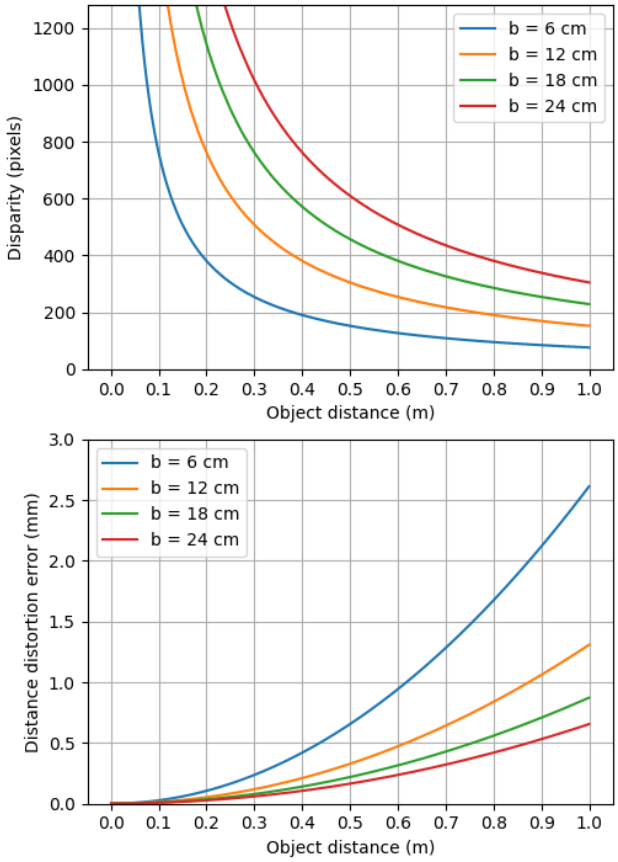



Figure 10. Baseline distance disparity and distortion error evaluation.

## IV. Conclusion and Future Work

The results showed a characterization process of an IoT stereo image sensor system, capable of capturing and transferring validated images via wireless commands from serial communication protocols, ready to be used in real agricultural field conditions. In this way, one of main contribution of this work is the construction of a system considering all parameters of casing, structure, power supply, communication, storage memory and hardware and software specifications, ready for use in a real field environment, while previous works only delved into software specifications for application in a controlled laboratory environment.

The developed system can be used in agricultural plantations, with a casing that protects the electrical components from sunlight, wind and light drizzle. It is necessary for a person to control the commands sent to the boards and help move the system, although the device can be adapted to be attached to a vehicle such as a tractor. With dedicated software

for identifying weed families, the device can then be used to generate detailed information, such as a distribution map of the occupancy of a given species in the cultivation area.

The MTF validation principles have demonstrated their importance in ensuring that captured images have sufficient contrast and are capable of observing details of plants with narrow and broad leaves, which allows the correct extraction of real-world data from the information generated by the sensors. Likewise, camera sensor distortions and 3D system calibration are essential so that the data can be used correctly.

Such developed embedded vision system can be useful for applications in 3D image processing, with several variable parameters that allow the adaptation of the system to different situations, although the power supply can be simplified to reduce the weight and power spent of the system, allowing the use of smaller batteries and fewer components (for example, with only a 12 Vdc to 5 Vdc converter and 9 W 12 V LED lamp).

For future steps, it is desired to carry out agricultural analyzes, considering weed families, as well as the inclusion of AI-based weed image process to identify plant species for agricultural control. In addition, an expansion of system's connectivity with other devices will also be realized.

## ACKNOWLEDGMENT

This work has been supported by the Brazilian Corporation for Agricultural Research (Embrapa) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

## REFERENCES

[1] B. M. Moreno and P. E. Cruvinel, "Characterization of an IoT Stereo Image Sensor System for Weed Control," in ALLSENSORS, International Conference on Advances in Sensors, Actuators, Metering and Sensing, 8th edition, pp. 1–7, 2023.

[2] V. R. Pathmudi, N. Khatri, S. Kumar, A. S. H. Abdul-Qawy and A. K. Vyas, "A systematic review of IoT technologies and their constituents for smart and sustainable agriculture applications," in Scientific African, Vol. 19, p. e01577, 2023.

[3] H. Abouziena and W. Haggag, "Weed control in clean agriculture: a review," Planta daninha, SciELO Brasil, Vol. 34, pp. 377-–392, 2016.

[4] S. C. Bhatla and M. A. Lal, "Plant physiology, development and metabolism," Springer, 2018.

[5] B. M. Moreno and P. E. Cruvinel, "Sensors-based stereo image system for precision control of weed in the agricultural industry," SENSORDEVICES 2018, The Ninth International Conference on Sensor Device Technologies and Applications, pp. 69–76, 2018.

[6] B. M. Moreno and P. E. Cruvinel, "Computer vision system for identifying on farming weed species," 2022 IEEE 16th International Conference on Semantic Computing (ICSC), USA, pp. 287–292, 2022.

[7] N. Higgs, B. Leyeza, J. Ubbens, J. Kocur, W. Kamp, T. Cory, C. Eynck, S. Vail, M. Eramian and I. Stavness, "ProTractor: a lightweight ground imaging and analysis system for early-season field phenotyping," in Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2629–2638, 2019.

[8] D. Li, L. Xu, X. Tang, S. Sun, X. Cai and P. Zhang, "3D imaging of greenhouse plants with an inexpensive binocular stereo vision system," in Remote Sensing, Vol. 9, pp. 508, 2017.

[9] M. I. Sadiq, S. M. P. Rahman, S. Kayes, A. H. Sumaita and N. A. Chisty, "A review on the imaging approaches in agriculture with crop and soil sensing methodologies," 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS), Morocco, pp. 1–7, 2021.

[10] M. Rajoriya and T. Usha, "Pattern recognition in agricultural areas," Journal of Critical Reviews, Vol. 7, pp. 1123–1127, 2020.

[11] J. Wäldchen, M. Rzanny, M. Seeland and P. Mäder, "Automated plant species identification—trends and future directions," PLoS computational biology, Vol. 14, pp. 1–19, 2018.

[12] M. P. Raj, P. R. Swaminarayan, J. R. Saini and D. K. Parmar, "Applications of pattern recognition algorithms in agriculture: a review," International Journal of Advanced Networking and Applications, Vol. 6, pp. 2495–2502, 2015.

[13] D. Mcloughlin, "Image processing apparatus for analysis of vegetation for weed control by identifying types of weeds," EP1000540, May 17, 2000.

[14] J. Gao and Z. Jin, "Bionic four-foot walking intelligent rotary tillage weeding device, has weeding system installed on back side of machine body, where gear in gear rotating mechanism transmits power to rotary shaft that is provided with weeding cutter," CN114794067, Jul 29, 2022.

[15] X. Jin, Y. Chen and J. Yu, "Precise weeding method for lawn and pasture based on cloud-killing spectrum, involves receiving images uploaded by each weeding robot, completing weed identification and outputting spraying instructions, and collecting and organizing massive weed data for big data applications," CN113349188, Sep 7, 2021.

[16] S. E. Mathe, M. Bandaru, H. K. Kondaveeti, S. Vappangi and G. S. Rao, "A survey of agriculture applications utilizing raspberry pi," 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, pp. 1–7, 2022.

[17] C. Balamurugan and R. Satheesh, "Development of raspberry pi and IoT based monitoring and controlling devices for agriculture," pp. 207–215, 2017.

[18] K. A. Patil and N. R. Kale, "A model for smart agriculture using IoT," 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, India, pp. 543–545, 2016.

[19] C. Bisdikian, "An overview of the Bluetooth wireless technology," in IEEE Communications Magazine, Vol. 39, pp. 86–94, 2001.

[20] Y. Fang, C. Tsai, J. MacDonald and Y. Pai, "Eliminating chromatic aberration in Gauss-type lens design using a novel genetic algorithm," in Applied Optics, Vol. 46, pp. 2401–2410, 2017.

[21] Y. Fang and C. Tsai, "Miniature lens design and optimization with liquid lens element via genetic algorithm," in Journal of Optics A: Pure and Applied Optics, Vol. 10, pp. 075304, 2008.

[22] C. C. Chen, C. M. Tsai and Y. C. Fang, "Optical Design of LCOS Optical Engine and Optimization With Genetic Algorithm," in Journal of Display Technology, Vol. 5, pp. 293–305, 2009.

[23] O. van Zwanenberg, S. Triantaphillidou, R. Jenkin and A. Psarrou, "Edge detection techniques for quantifying spatial imaging system performance and image quality," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1871–1879, 2019.

[24] N. Kawagishi, R. Kakinuma and H. Yamamoto, "Aerial image resolution measurement based on the slanted knife edge method," in Optics Express Vol. 28, pp. 35518–35527, 2020.

[25] A. A. Naumov, A. V. Gorevoy, A. S. Machikhin, V. I. Batshev and V. E. Pozha, "Estimating the quality of stereoscopic endoscopic systems," in Journal of Physics: Conference Series, Vol. 1421, pp. 012044, 2019.

[26] J. L. Alió, P. Schimchak, R. Montés-Micó and A. Galal, "Retinal image quality after microincision intraocular lens implantation," in Journal of Cataract & Refractive Surgery, Vol. 31, pp. 1557–1560, 2005.

[27] E. Oh and J.-K. Choi, "GOCI image enhancement using an MTF compensation technique for coastal water applications," in Opt. Express, Vol. 22, pp. 26908–26918, 2014.

[28] L. Yang, B. Wang, R. Zhang, H. Zhou and R. Wang, "Analysis on location accuracy for the binocular stereo vision system," in IEEE Photonics Journal, Vol. 10, no. 1, pp. 1–16, Art no. 7800316, Feb. 2018.

[29] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge: Cambridge University Press, 2004.

[30] Y. Wang, X. Wang, Z. Wan, and J. Zhang, "A method for extrinsic parameter calibration of rotating binocular stereo vision using a single feature point," in Sensors, Vol. 18, Art no. 3666, pp. 1–16, 2018.

[31] J. Salvi, X. Armangué and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," in Pattern Recognition, Vol. 35, Issue 7, pp. 1617–1635, 2002.

[32] Y. M. Wang, Y. Li and J. B. Zheng, "A camera calibration technique based on OpenCV," The 3rd International Conference on Information Sciences and Interaction Sciences, Chengdu, China, pp. 403–406, 2010.

[33] S. Yang, Y. Gao, Z. Liu and G. Zhang, "A calibration method for binocular stereo vision sensor with short-baseline based on 3D flexible

control field," in Optics and Lasers in Engineering, Vol. 124, pp. 105817, 2020.

[34] J. Sun, X. Chen, Z. Gong, Z. Liu and Y. Zhao, "Accurate camera calibration with distortion models using sphere images," in Optics & Laser Technology, Vol. 65, pp. 83–87, 2015.