SN Computer Science (2024) 5:582
https://doi.org/10.1007/s42979-024-02849-7

ORIGINAL RESEARCH

A Novel Multi-camera Fusion Approach at Plant Scale: From 2D to 3D

Edgar S. Correa[1,2,3] · Francisco C. Calderon[1] · Julian D. Colorado[1,4]

Received: 10 March 2022 / Accepted: 30 March 2024

Abstract

Non-invasive crop phenotyping is essential for crop modeling, which relies on image processing techniques. This research presents a plant-scale vision system that can acquire multispectral plant data in agricultural fields. This paper proposes a sensory fusion method that uses three cameras, Two multispectral and a RGB depth camera. The sensory fusion method applies pattern recognition and statistical optimization to produce a single multispectral 3D image that combines thermal and near-infrared (NIR) images from crops. A multi-camera sensory fusion method incorporates five multispectral bands: three from the visible range and two from the non-visible range, namely NIR and mid-infrared. The object recognition method examines about 7000 features in each image and runs only once during calibration. The outcome of the sensory fusion process is a homographic transformation model that integrates multispectral and RGB data into a coherent 3D representation. This approach can handle occlusions, allowing an accurate extraction of crop features. The result is a 3D point cloud that contains thermal and NIR multispectral data that were initially obtained separately in 2D.

Keywords Multi-spectral imagery · Light-field plenoptic cameras · Phenotyping · Plant modeling · 3D plant morphology

Introduction

The world population and food demand are increasing, making the development of sustainable agricultural technologies a vital task [1]. Rice is one of the most important foods worldwide. Experimental phenotyping of different rice varieties enables genomic selection models and assessment of agronomic traits such as temperature and humidity tolerance, radiation levels, aluminum toxicity in soils, and biotic stress [2–6]. Phenotypic quantification requires accurate morphological modeling, which offers useful information to validate new agricultural varieties for higher productivity and food security [7, 8]. Plant morphological traits are key variables in estimating grain yield and crop health. Traditional methods are often invasive [9] or destructive [10], depending on biological samples [11–14]. To overcome the drawbacks of traditional methods, image processing techniques have emerged as a non-destructive alternative. These techniques allow qualitative and quantitative analysis of light absorption and reflection at different bands, enabling the characterization of crop conditions. This, for instance, permits the detection of nitrogen-deficient plants [15–17].

Abiotic stress in plants causes changes in fluorescence due to the absorption and reflection of light at different bands. These variations occur within the 650 to 800 nm range of the electromagnetic spectrum, corresponding to the chlorophyll fluorescence [18, 19]. Traditional methods usually involve direct point measurements, using two main components: (i) image data captured by RGB or multispectral cameras, such as near-infrared, mid-infrared, or thermal cameras, and (ii) three-dimensional sensors, such as LiDAR, stereo cameras, or plenoptic cameras. The fusion

This article is part of the topical collection "Advances in Applied Image Processing and Pattern Recognition" guest edited by K C Santosh.

✉ Edgar S. Correa
e_correa@javeriana.edu.co

Francisco C. Calderon
calderonf@javeriana.edu.co

Julian D. Colorado
coloradoj@javeriana.edu.co

[1] School of Engineering, Pontificia Universidad Javeriana, Bogotá, Cra. 7 No. 40-62, 110311, Bogotá, Colombia

[2] Faculty of Sciences, Université de Montpellier, Montpellier, France

[3] CIRAD, AGAP, Montpellier, France

[4] Omics Science Research Institute, iOMICAS, Pontificia Universidad Javeriana, Cali 760031, Colombia

Published online: 23 May 2024

of data from these components allows the generation of a four-dimensional (4D) model [20].

To the best of the authors' knowledge, few works in the literature have used light-field cameras to reconstruct 4D plant models. In this arena, the PhenoBot research is one of the few studies that address plant phenotyping through 4D models using a plenoptic camera [21]. Progress in this area has been restricted to using a single frame or a single camera to extract plant features, despite the informative data of the plants being available in different multispectral cameras. This requires a multi-camera fusion method. To tackle this challenge, our research aimed to develop a method for extracting non-invasive multispectral data of the plant through a multi-camera fusion method, enabling the acquisition of infrared and thermal images in 3D space that were initially in 2D.

Related Work

Multi-spectral sensory fusion enhances robustness and reliability across a broader range of applications compared to using only single-wavelength information. Convolutional Neural Network (CNN) is a popular deep network architecture widely employed for analyzing visual imagery [22]. Some studies have introduced CNNs for multispectral remote sensing image analysis to enhance the performance of detection algorithms. These CNN-based detectors are trained on large-scale satellite image datasets [23–25]. While these studies focus on object detection applications, they do not encompass homographic transformation to unify images from different sensors into a single image, they lack a multi-camera fusion approach. Furthermore, they typically require a large number of images for training, often exceeding 1200 images. Although these studies introduce object detection applications utilizing multispectral information, they do not employ a multisensory fusion approach [26]. Other research integrates 3D information with multispectral remote sensing images to model 3D tree canopies [27, 28], achieving sensory fusion through public libraries and free software. However, these approaches do not utilize sensory fusion via homographic transformations. This is because the distance from the sensor to the canopy is often significant, making such transformations unnecessary. Conversely, our research focuses on sensors positioned close to the plants, necessitating homographic transformations. Another approach involves multi-sensory fusion applications over different point clouds using 3D spatial join techniques [13, 29, 30]. While promising, this approach is not suitable when the multispectral information is obtained from 2D sensors. Lastly, Multi-Camera Sensor Fusion is commonly utilized for visual odometry using artificial neural networks [31], or tracking interesting objects, but it typically does not integrate multispectral information [32, 33].

This research presents a multisensory fusion at plant scale. The challenge in this proposal is to adapt techniques traditionally associated with other contexts [34–36]. By utilizing both 2D and 3D cameras, this study proposes a multisensory data approach. Accordingly, our research develops a strategy based on pattern recognition and statistical optimization to model projective transformations through a complex object recognition application [37]. The visible light spectrum (VIS) wavelengths are captured using the Plenoptic Raytrix R42 camera and the Kinect One V02 sensor. The non-visible near-infrared spectrum (NIR) is captured using the Parrot Sequoia camera, while the mid-infrared spectral band (MIR) is captured using the Fluke Ti400 thermal camera.

Section "Materials and Setup" presents the materials, which encompass the mechanical configuration and calibration of the cameras. Section "Methodology" introduces the methodology, covering (i) feature detection, (ii) feature matching with statistical optimization, (iii) homographic model transformation, and the integration with multispectral cameras. Finally, the results and conclusions are presented in Chapters 4 and 5, respectively.

Materials and Setup

Images Acquisition

The sensory fusion approach is implemented using three cameras: (i) a 3D camera operating in the visible spectrum (VIS), (ii) a near-infrared (NIR) multispectral camera, and (iii) a mid-infrared thermal multispectral camera. This approach requires the setup of the mechanical assembly and the configuration of acquisition software.

Camera Assembly and Mounting Structure

In Fig. 1, two different configurations are observed. These camera setups share the characteristic of integrating a 3D camera with three VIS channels, a multispectral camera with an infrared channel, and a thermal camera with only one channel. The entire assembly is mounted on a tripod to ensure stability and maintain the integrity of the homographic alignment.

Plenoptic Camera Calibration

Light field cameras have garnered attention for their innovative capabilities. This technology captures both the intensity and direction of light rays as they propagate through space.

MLA Micro-Lens Array calibration Figure 2a illustrates the structure of the plenoptic sensor developed by Raytrix. This image highlights the need for camera calibration due to

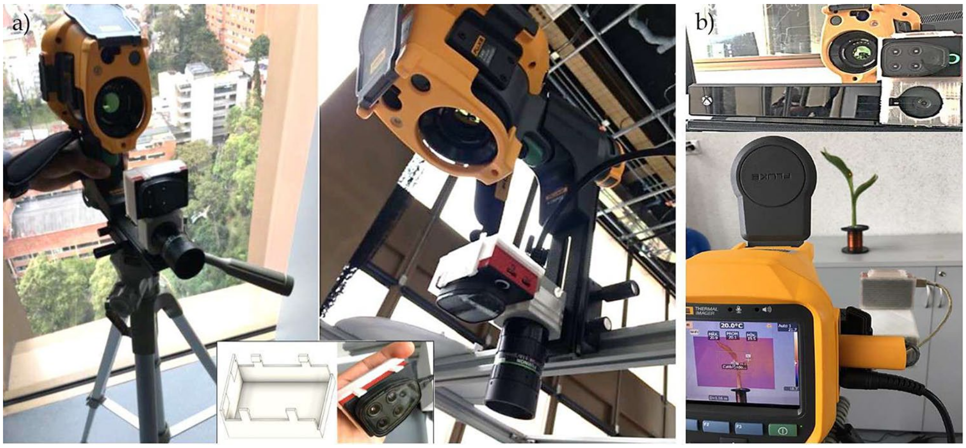


Fig. 1 **a** Mechanical assembly of plenoptic camera, multi-spectral camera, and thermal camera. **b** Mechanical assembly of Kinect V2, multi-spectral camera, and thermal camera

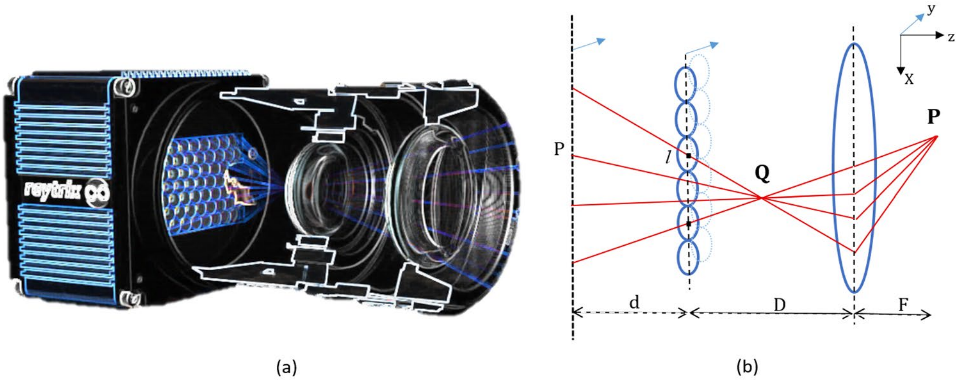


Fig. 2 Plenoptic camera. **a** 3D light field camera [9]. **b** Projective model of Plenoptic camera based on the micro-lenses array [40]

the micro-lens architecture. In Fig. 2b, the general projective model of a Plenoptic camera, based on literature [38, 39], is presented. In this model, point P represents the real spatial information of the scene (P_x, P_y, P_z), and the light rays from point P are captured by the main lens, resulting in corresponding points Q forming the image captured by the camera. The set of micro-lenses, denoted as l, forms the basis of plenoptic technology and directly influences the generation of pixels p. When $P_z > 0$, Eq. 1 describes the relationship, where d is the distance between the camera sensor and the micro-lens array, D is the distance between the micro-lens array and the main lens, and F is the focal length of the main lens and the object in the scene [38].

$$\frac{1}{F}=\frac{1}{P_z}-\frac{1}{Q_z} \tag{1}$$

The camera calibration process is performed using the RxLive tool and comprises two essential components: the calibration filter and a light source, as demonstrated in Fig. 3.

Three crucial components are essential to the calibration process, necessitating manual fine-tuning: the camera's primary lens (in this case, a 12 mm lens), the diaphragm aperture (controlling the light influx to the sensor), and the focus setting (establishing the camera-to-subject distance). Alterations to any of these parameters necessitate a subsequent recalibration. As depicted in Fig. 3, the distance from the camera to the desktop is recorded at 360 mm, suggesting the need for corresponding adjustments to the focal length.

Image illumination is controlled by adjusting the exposure time while maintaining a constant aperture setting. For the given lighting conditions, the exposure time is established at 55 milliseconds. In Fig. 4a, an overexposure effect is evident, a common occurrence when the light source is positioned directly in front of the camera, as noted by Co et al. (2022). Figure 4b showcases an image with optimal exposure, whereas Fig. 4c illustrates the calibration process for the micro-lens array. Images obtained using the calibrated camera setup are presented in Fig. 5a.

The metric Calibration is performed using the RxLive 5.0 software calibration wizard. A 22 mm calibration target is utilized, and a total of 44 images are captured with varying positions, inclinations, and rotations. Figure 6a depicts the calibration interface, while Fig. 6b illustrates the 3D acquisition process.

Kinect: 3D sensor acquisition. The acquisition of data from this sensor is performed using the MATLAB

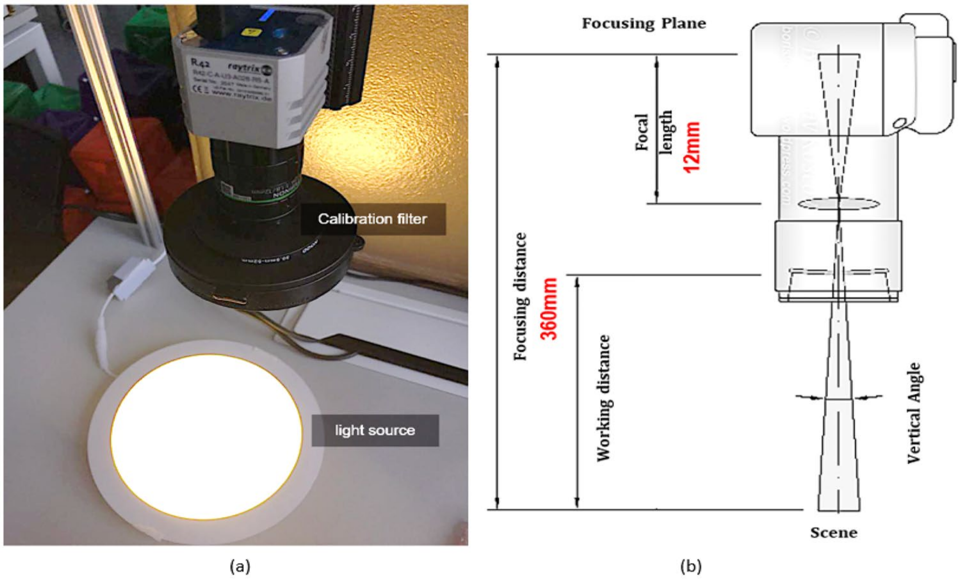


Fig. 3 Camera Ratrix R42 with filter calibration disk and light source [40]

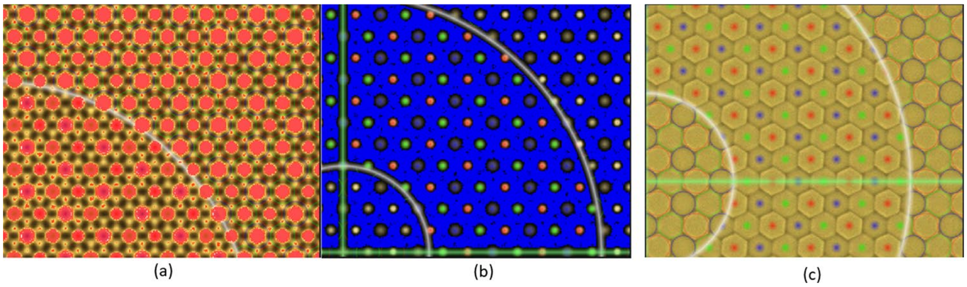


Fig. 4 **a** Conditions of over-exposure. **b** Good lighting conditions. **c** Micro Lens Array calibration [40]

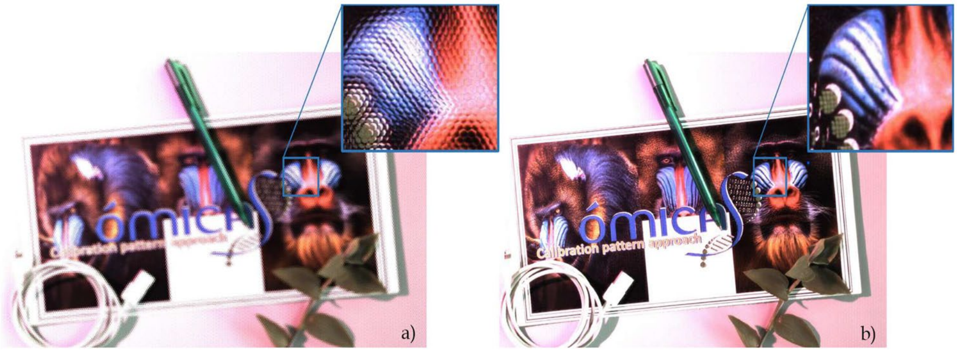


Fig. 5 **a** Light field image captured with a R42 Raytrix plenoptic camera, at a focal length of 360 mm. **b** Calibrated Light field image captured with a R42 Raytrix plenoptic camera, at a focal length of 360 mm

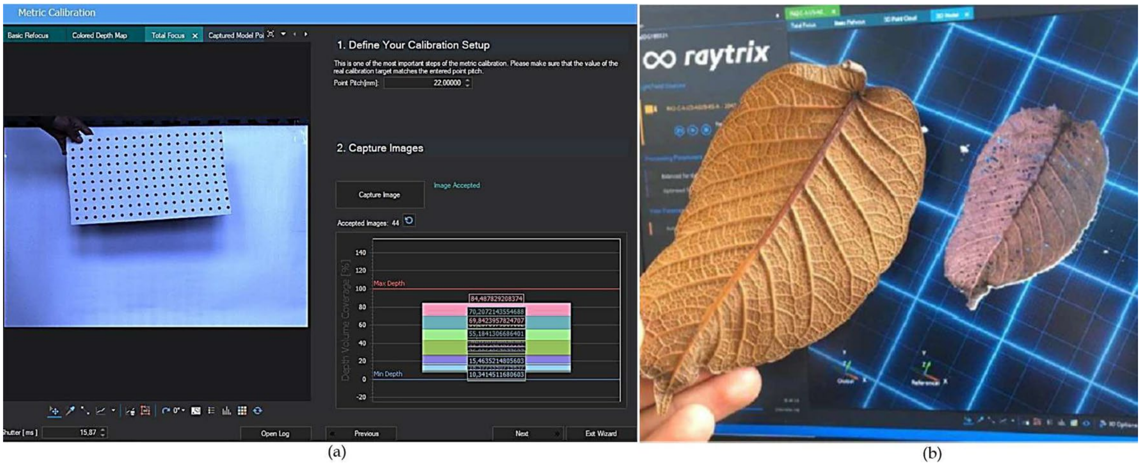
programming tool. It is important to ensure that the Kinect SDK and Runtime drivers are installed.

Parrot Sequoia: Multi-spectral camera acquisition. The configuration of this sensor is carried out using the service provided by the camera via a WiFi connection.

Fluke ti400: Thermal image acquisition. The data captured by this camera is stored on a USB memory device, which is then read and processed using the SmartView software tool.

Methodology

The use of 2D and 3D cameras proposes a multisensory data approach. Thereby, this research work develops a strategy based on pattern recognition and statistical optimization to model projective transformation through a complex object recognition application.



1. Define Your Calibration Setup

This is one of the most important steps of the metric calibration. Please make sure that the value of the real calibration target matches the entered point pitch.

Point Pitch[mm]: 22,00000

2. Capture Images

Accepted Images: 44

Shutter [ms] 15,87

Fig. 6 **a** Metric calibration interface of RxLive 5.0. **b** 3D acquisition with Plenoptic Camera in RxLive 5.0 Software

Complex Object Detection Approach

The methodology is elaborated in four stages, as depicted in Fig. 7. The first stage involves image acquisition with calibrated cameras, followed by feature detection using descriptor vectors in the second stage. The third stage encompasses the matching of these features based on probabilistic optimization. Finally, the fourth stage involves estimating the spatial transformation model and implementing homographic projection to validate the methodology.

Figure 8 shows the experiment structure. It consists of having the interest object on the right, a complex image in shape and color distribution, on the left, the scene desired to detect this object.

Feature Detection

To accurately characterize the scene, prominent and distinctive areas of the image must be detected. The robustness of the object detector application relies on the descriptor used. Features should be invariant to illumination, 3D projective transforms, and common object variations. In this research, the Scale Invariant Feature Transform (SIFT) approach is employed.

SIFT transforms an image into an extensive collection of local feature vectors. The algorithm is inspired by the response of neurons in the inferior temporal cortex of primates to vision [41]. The resulting feature vector comprises 128 dimensions, with each descriptor assigned a position in the image denoted by coordinates (X, Y). A more complex image, containing a greater number of details, will yield a larger number of descriptors.

Figure 9 illustrates the position (x, y) of each descriptor found in the image, marked with an asterisk.

Algorithm 1 Feature matching by force.

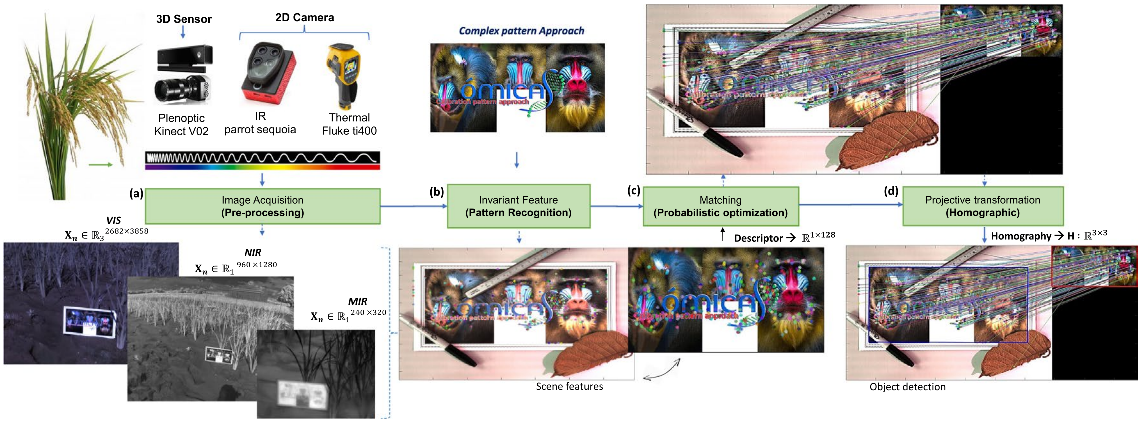

3D Sensor
2D Camera
Plenoptic
Kinect V02
IR
parrot sequoia
Thermal
Fluke ti400
Complex pattern Approach
(a)
Image Acquisition
(Pre-processing)
(b)
Invariant Feature
(Pattern Recognition)
(c)
Matching
(Probabilistic optimization)
(d)
Projective transformation
(Homographic)
VIS
$X_n \in \mathbb{R}_3^{2682\times3858}$
NIR
$X_n \in \mathbb{R}_1^{960\times1280}$
MIR
$X_n \in \mathbb{R}_1^{240\times320}$
Descriptor → $\mathbb{R}^{1\times128}$
Homography → H : $\mathbb{R}^{3\times3}$
Scene features
Object detection

Fig. 7 Object detection approach. **a** Image acquisition. **b** Pattern recognition stage. **c** Matching–Optimization. **d** Projective transformation

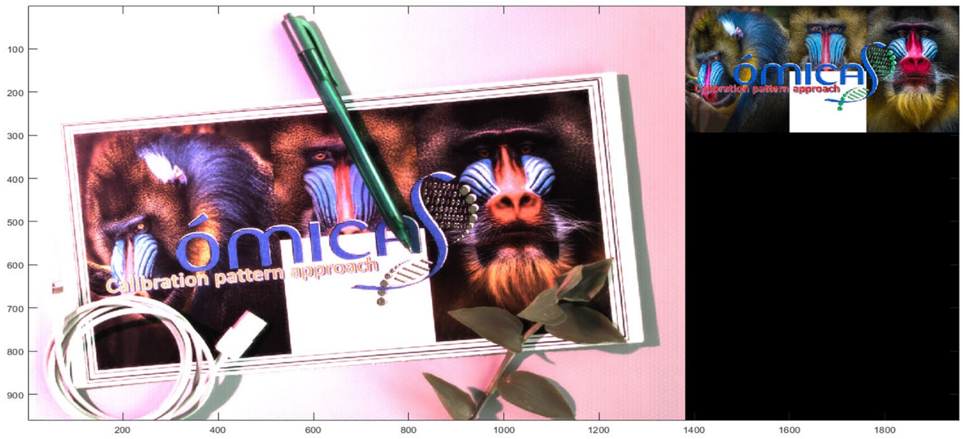


Fig. 8 Topology of the experiment design, on the right the object of interest, and on the left, a scene containing several objects with many similar characteristics

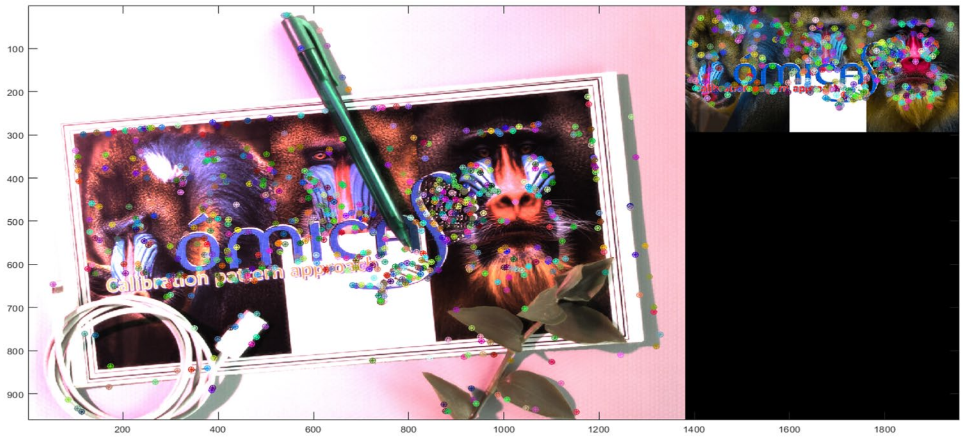


Fig. 9 The composition of images, indicating with an asterisk marks the position of the descriptors generated with SIFT algorithm for each image

```
for $i = 1 : size(A)$ do
    $X \Leftarrow cos^{-}1(A(i) * B)$
    $Sort(X)$
    if $Match(1) < distRatio * Match(2)$ then
        $MatchTable(i) \Leftarrow Match(1)$
    elseState $MatchTable(i) \Leftarrow 0$
    end if
end for
```

Feature Matching

The matching stage establishes the relationship between the information generated in two images. This correspondence links each feature by calculating the distance metric between descriptors. In the scene, there are 7100 features, while the interest object image contains 1184 features. Consequently, Fig. 10 displays 7100 matches generated.

To optimize the pattern recognition application, it is desirable to process the least amount of information possible. Algorithm 1 presents the feature correspondence by brute force, achieved through the security metric *distRatio* to identify the most prominent matches. Figure 11 illustrates the result of the brute force matching filter applied to Fig. 10, yielding 251 matches.

Transform Model Estimation

The transformation model is a mapping function that establishes a relationship between the object of interest in the scene and the reference image, which solely features the object of interest in the foreground. This mapping is accomplished through the homography matrix, calculated using the positional information of the descriptors. To achieve this, the Random Sample Consensus (RANSAC) method is employed as a search strategy. RANSAC is an iterative technique used to estimate the parameters of a mathematical model from a set of observed data, which may include both inliers and outliers. The methodology for implementing RANSAC is outlined in Algorithm 2.

Algorithm 2 RANSAC Algorithm

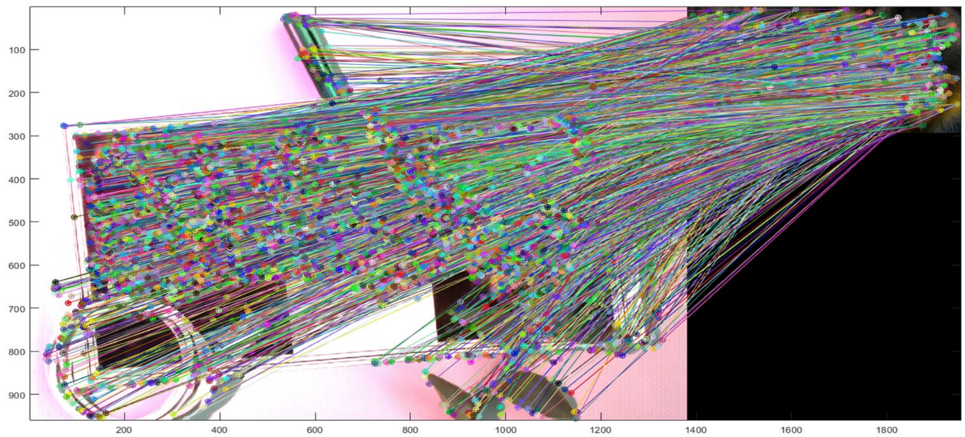


Fig. 10 7100 matches between referenced image and scene image

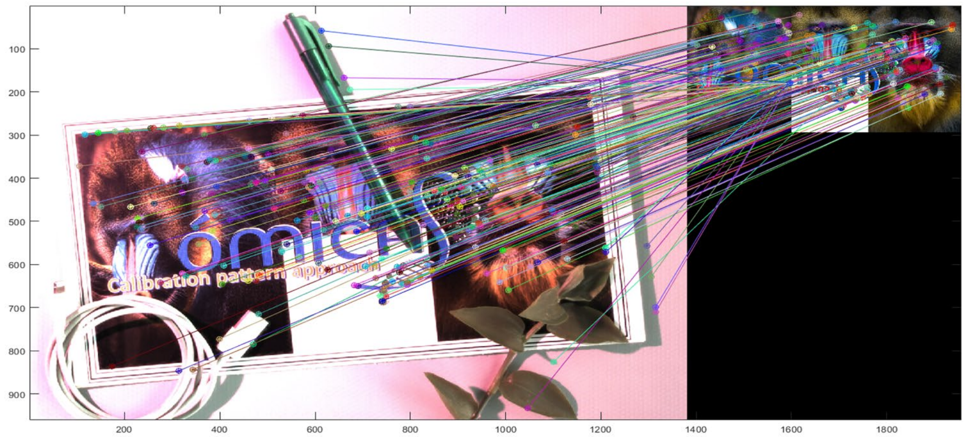


Fig. 11 251 matches generated with the algorithm 1 over 7100 matches from Fig. 10

```
$max\ number\ of\ Inlier \leftarrow 0$
Calculate maximum iteration
**while** $number\ of\ Iterations\ <\ N$ **do**
    **I. Hypothesis generation**
    $S_i \leftarrow select\ minimal\ subset\ of\ samples\ randomly$
    $M_i \leftarrow generate\ a\ hypothesis\ (model)\ from\ s_i$
    **II. Hypothesis evaluation**
    $Calculate\ error\ from\ estimated\ model$
    $I_i \leftarrow count\ the\ number\ of\ detected\ inliers$
    **if** $I_i\ >\ max\ number\ of\ Inlier$ **then**
        $Update\ max\ number\ of\ Inlier$
        $UpdateN$
    **end if**
    $number\ of\ Iteration\ \leftarrow number\ of\ Iteration + 1$
**end while**
```

This process involves estimating the optimal transformation statistically, based on a chi-square probability distribution. The probability that a point is an inlier is set to $\alpha = 0.95$, and to calculate the homography, $\sigma^2 = 5.99$ [42].

This approach ensures that the number of samples chosen is representative, guaranteeing with a probability p that at least one of the random samples of points s is free of outliers, meaning the estimated transformation is free of outliers with a probability of $p = 0.99$.

To ensure this, the probability of any selected data being an inlier is defined as ϵ, and $1 - \epsilon$ as the probability of selecting an outlier. At least N selections of s points are required to ensure $(1 - \epsilon^s)^N = 1 - p$, resulting in the model represented by Eq. 2 with ϵ in Eq. 3.

$$N = \frac{log(1 - p)}{log(1 - (1 - \epsilon^s))} \tag{2}$$

$$\epsilon = \frac{1 - number\ of\ inlier}{Total\ number\ of\ points} \tag{3}$$

The consensus process concludes when the modeled probability exceeds the threshold set by the number of events. The spatial transformation is accomplished using the homographic matrix described in Eqs. 4 to 6. The first equation relates to a rotational transformation, the second incorporates linear transformations along the (x, y) axes, and the third represents a complete homography transformation in space. The latter transformation, expressed in Eq. 7, is utilized in this work.

$$\begin{vmatrix} X_T \\ Y_T \\ 1 \end{vmatrix} = \begin{vmatrix} Cos(\theta) & -Sin(\theta) & 0 \\ Sin(\theta) & Cos(\theta) & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{4}$$

$$\begin{vmatrix} X_T \\ Y_T \\ 1 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{5}$$

$$\begin{vmatrix} X_T \\ Y_T \\ 1 \end{vmatrix} = \begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{6}$$

$$X_T = H.X_R \tag{7}$$

The goal is to find the transformation matrix H, defined as: h = $\begin{vmatrix} h_{11} & h_{12} & h_{13} & h_{21} & h_{22} & h_{23} & h_{31} & h_{32} & h_{33} \end{vmatrix}$. The transformation $X_T = H \cdot X_R$ can be expressed as a linear system $Ah = 0$ [43–45]. This system is solved using Gaussian elimination with a pseudo-inverse method, as shown in Eq. 8. The matrix resolution is implemented with matrices A and B, which are presented in Eqs. 9 and 10 respectively.

$$h = (A'^T.A')^{-1} A'^T.b' \tag{8}$$

$$A = \begin{vmatrix} X_{R1} & Y_{R1} & 1 & 0 & 0 & 0 & -X_{R1}.X_{T1} & -Y_{R1}.X_{T1} & -X_{T1} \\ 0 & 0 & 0 & X_{R1} & Y_{R1} & 1 & -X_{R1}.Y_{T1} & -Y_{R1}.y_{T1} & -y_{T1} \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ X_{Rn} & x_{Rn} & 1 & 0 & 0 & 0 & -X_{Rn}.X_{Tn} & -Y_{Rn}.X_{Tn} & -X_{Tn} \\ 0 & 0 & 0 & X_{Rn} & Y_{Rn} & 1 & -X_{Rn}.Y_{Tn} & -Y_{Rn}.Y_{Tn} & -Y_{Tn} \end{vmatrix} \tag{9}$$

$$b = \begin{vmatrix} X_{T1} \\ Y_{T1} \\ . \\ . \\ X_{Tn} \\ Y_{Tn} \end{vmatrix} = \begin{vmatrix} X_{R1} & Y_{R1} & 1 & 0 & 0 & 0 & -X_{R1}.X_{T1} & -Y_{R1}.X_{T1} & -X_{T1} \\ 0 & 0 & 0 & X_{R1} & Y_{R1} & 1 & -X_{R1}.Y_{T1} & -Y_{R1}.y_{T1} & -y_{T1} \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ X_{Rn} & x_{Rn} & 1 & 0 & 0 & 0 & -X_{Rn}.X_{Tn} & -Y_{Rn}.X_{Tn} & -X_{Tn} \\ 0 & 0 & 0 & X_{Rn} & Y_{Rn} & 1 & -X_{Rn}.Y_{Tn} & -Y_{Rn}.Y_{Tn} & -Y_{Tn} \end{vmatrix} \cdot \begin{vmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{vmatrix} \tag{10}$$

Image Transformation

In Fig. 12, the reference image is depicted in red, while the same object is shown in blue in the scene image. The transformation is performed using the homographic matrix ’h’, which is applied to each corner of the complex object.

The output of the complex object recognition stage is the homographic matrix [3×3]. This matrix models the projective transformation of the scenes, aligning two images from different cameras to create a single multispectral image. This relationship enables the development of a general application of sensory fusion for multispectral images.

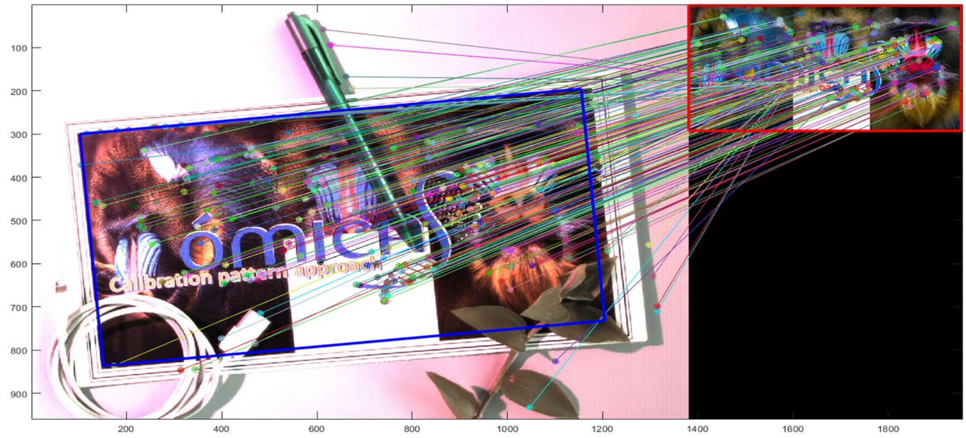


Fig. 12 Object recognition with occlusion under controlled conditions

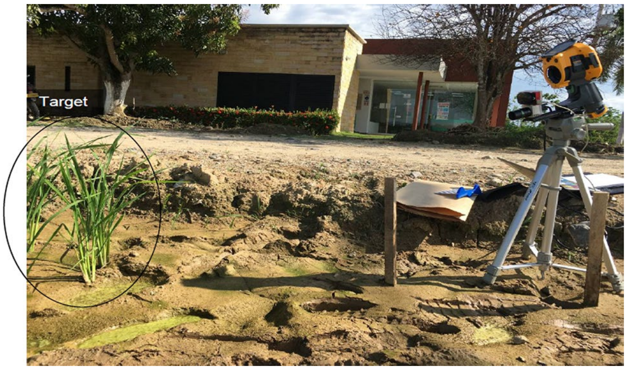


Fig. 13 Plant image acquisition, in filed conditions, whit plenoptic, multispectral and thermal cameras configuration

Multi-Sensory Image Making Up

Each of the cameras integrated into this research produces 2D information. The first step towards achieving sensory fusion is to utilize this 2D information for pattern recognition. Subsequently, the resulting homography matrices are used to relate the 2D information to the referenced 3D sensor. The research revolves around the acquisition process depicted in Fig. 13.

In Fig. 14, the images acquired by each sensor for the same scene are depicted: the plenoptic, multispectral, and thermal cameras, respectively. It is evident that each camera has a different resolution and covers a distinct area of the crop. Furthermore, the topology of each image varies, even though they were captured simultaneously from the same scene. Specifically, the thermal image appears smaller relative to the other two.

The challenge lies in developing an algorithm capable of generating a single composite image by transforming each available channel. These channels consist of (i) RGB - three channels from the 3D sensor, (ii) Near IR - one channel from the infrared image, and (iii) Medium IR - one channel from the thermal image.

The objective is to align the multispectral information captured by each camera with the reference frame of the 3D sensor. This involves two steps: (i) establishing the relationship between the plenoptic and multispectral NIR camera, and (ii) determining the relationship between the plenoptic and multispectral MIR camera.

Matching between Plenoptica camera and NIR camera

Figure 15a illustrates the homographic relationship between the plenoptic camera and the multispectral NIR camera. The region of the image captured with the plenoptic camera is highlighted in red, while the region captured with the multispectral camera is indicated in cyan. The homographic transformation obtained using Eq. 8 is depicted in blue.

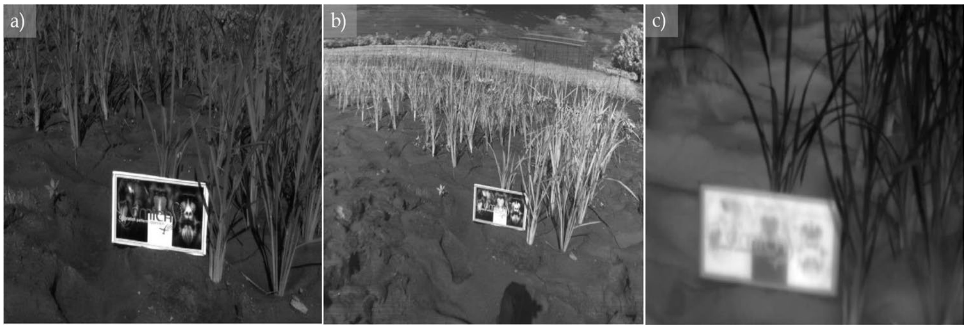


Fig. 14 **a** Plenoptic image with raytrix R42 camera. Resolution = [960×1381]. **b** Infrared IR image with parrot sequoia camera. Resolution = [960× 1280]. **c** Thermal IR image with t400 fluke camera. Resolution = [240×320]

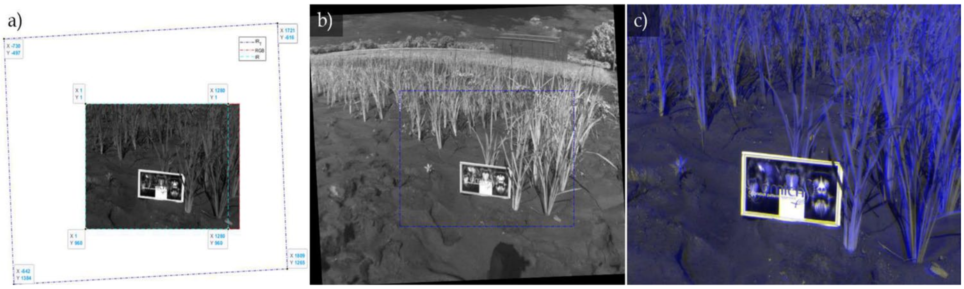


Fig. 15 **a** Homographic projection between the plenoptic camera and the NIR multispectral camera. **b** Homographic transformation of the NIR multispectral image. **c** RGN Image. Composed of the RG channels of the plenoptic image and the NIR channel of the multispectral camera. Resolution = [960×1381]

Figure 15b depicts the transformation of the entire NIR multispectral image. The blue box represents the overlapping region of the multispectral camera with the frame of the plenoptic image, providing a new channel that can be related to the three-dimensional information. In Fig. 15c, an image composed of the RG channels of the plenoptic camera and the NIR multispectral channel is presented.

Matching between Plenoptic camera and thermal camera

Figure 16a illustrates the homographic relationship between the plenoptic camera and the thermal multispectral camera MIR. The region of the image captured with the plenoptic camera is highlighted in red, while the region captured with the thermal multispectral camera is indicated in cyan. The homographic transformation obtained using Eq. 8 in the thermal image is depicted in blue.

Figure 16b illustrates the transformation of the complete thermal multispectral image. In contrast to Fig. 15b, the entire image is utilized in this case, and there is no need to crop a fraction.

In Fig. 16c, an image is composed using the RG channels of the plenoptic camera and the thermal MIR multispectral channel.

Figure 17a presents an image composition consisting of the R channel of the plenoptic camera, the IR channel of the NIR multispectral camera, and the IR channel of the MIR thermal multispectral camera. Finally, Fig. 17b displays the five channels available for generating different image compositions and calculating the vegetative index.

Figure 18 displays the homographic matrices that establish the correspondence between common features in each scene, such as the position of the known pattern. This information facilitates the integration of images into a single frame, even when they are captured by different cameras. In this application, the target frame is the one associated with three-dimensional information, which in this research is linked to the plenoptic camera or Kinect sensor.

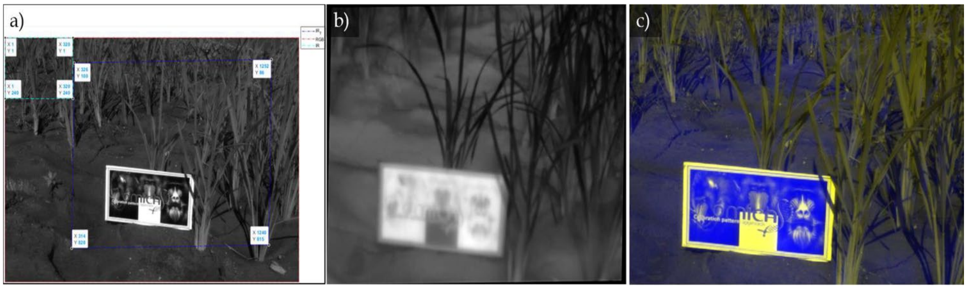


Fig. 16 **a** Homographic projection between the plenoptic camera and the thermal multispectral camera MIR. **b** Homographic transformation of the thermal multispectral MIR image. **c** RGN2 image. Composed of the RG channels of the plenoptic image and the MIR channel of the thermal multispectral camera. Resolution = [716×915]

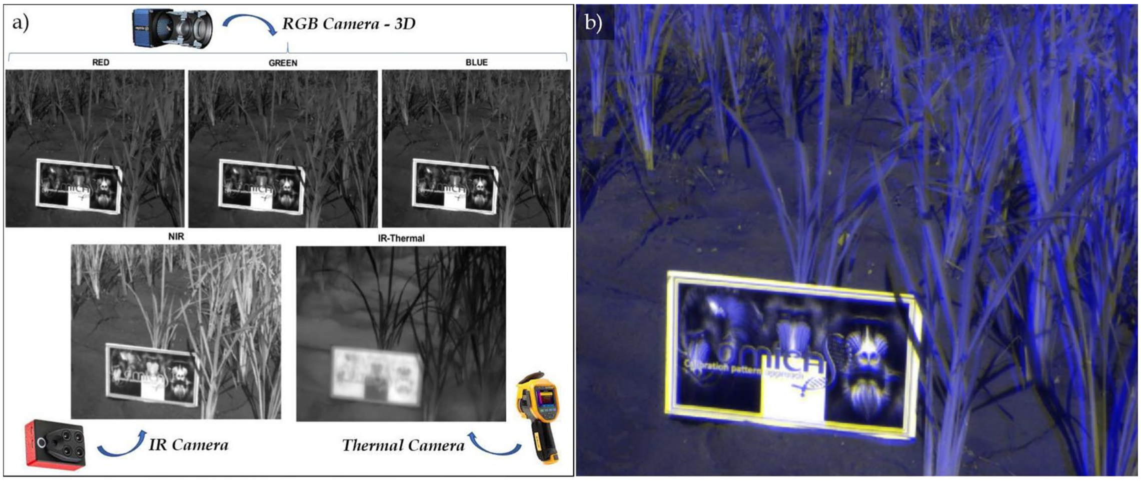


Fig. 17 **a** Five channels in the same frame, RGB from the plenoptic camera, NIR from the multispectral camera, and IR-Thermal from the thermal camera. Resolution = [960×1381].**b** RNN image. Composed of the R channel of the plenoptic image, the IR channel of the NIR multispectral camera and the IR channel of the MIR thermal mulltiespectral camera. Resolution = [716×915]

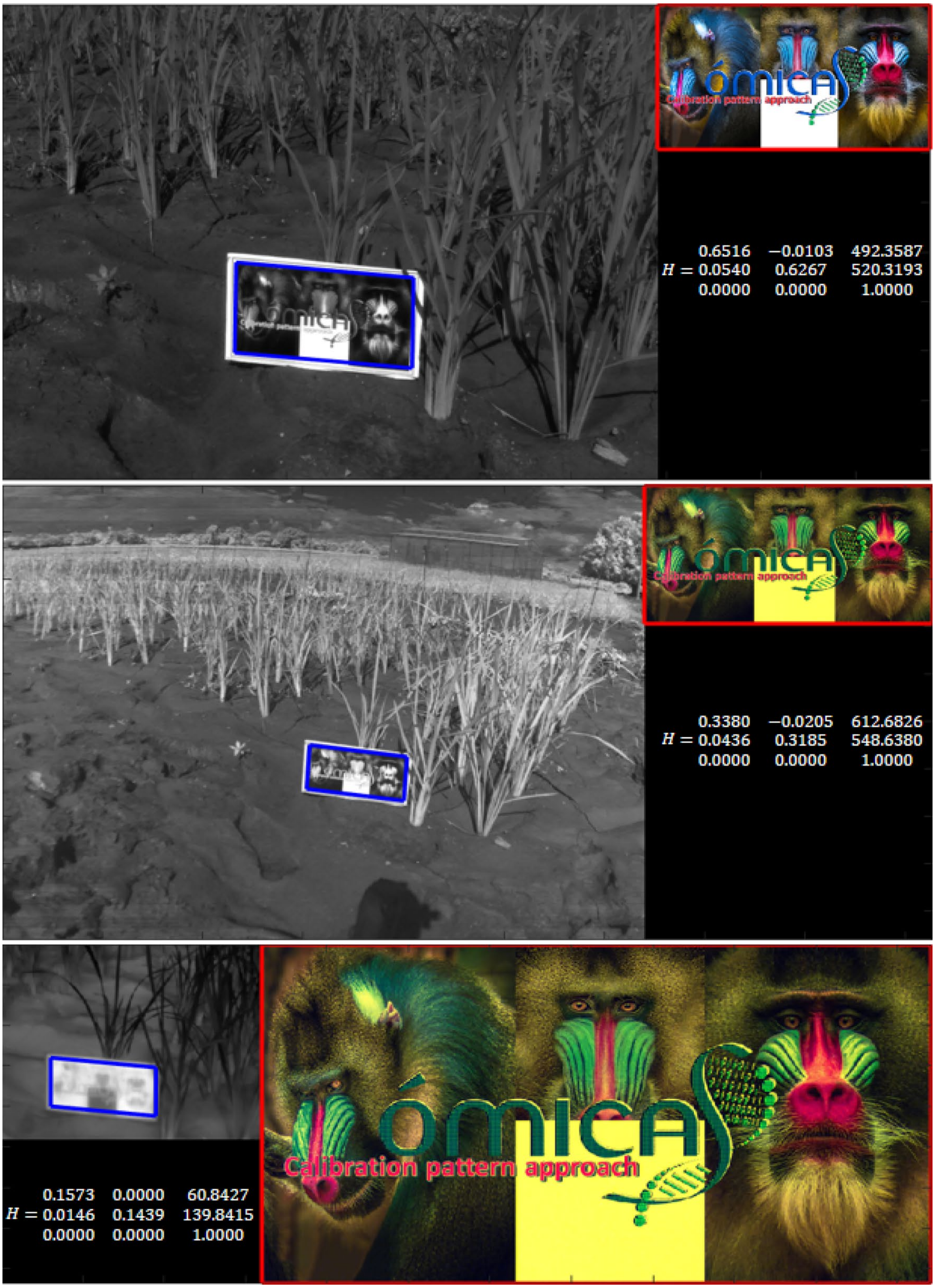


Fig. 18 Homographic relationship through complex pattern visualization changes in shape and color

Proposed System Model

Figure 19 illustrates the diagram depicting the stages involved in implementing the multispectral sensory fusion approach. The diagram is structured into three modules: (a) The first module involves object recognition with a complex pattern within the same scene captured by three different sensors. (b) The second module focuses on setting up an image with various configurations of channels from the three introduced cameras, integrating them with 3D spatial information. (c) The final module aims to validate the model characterized by homographic projections in a new scene. This validation stage solely considers the homographic model with new images, without taking into account the complex pattern.

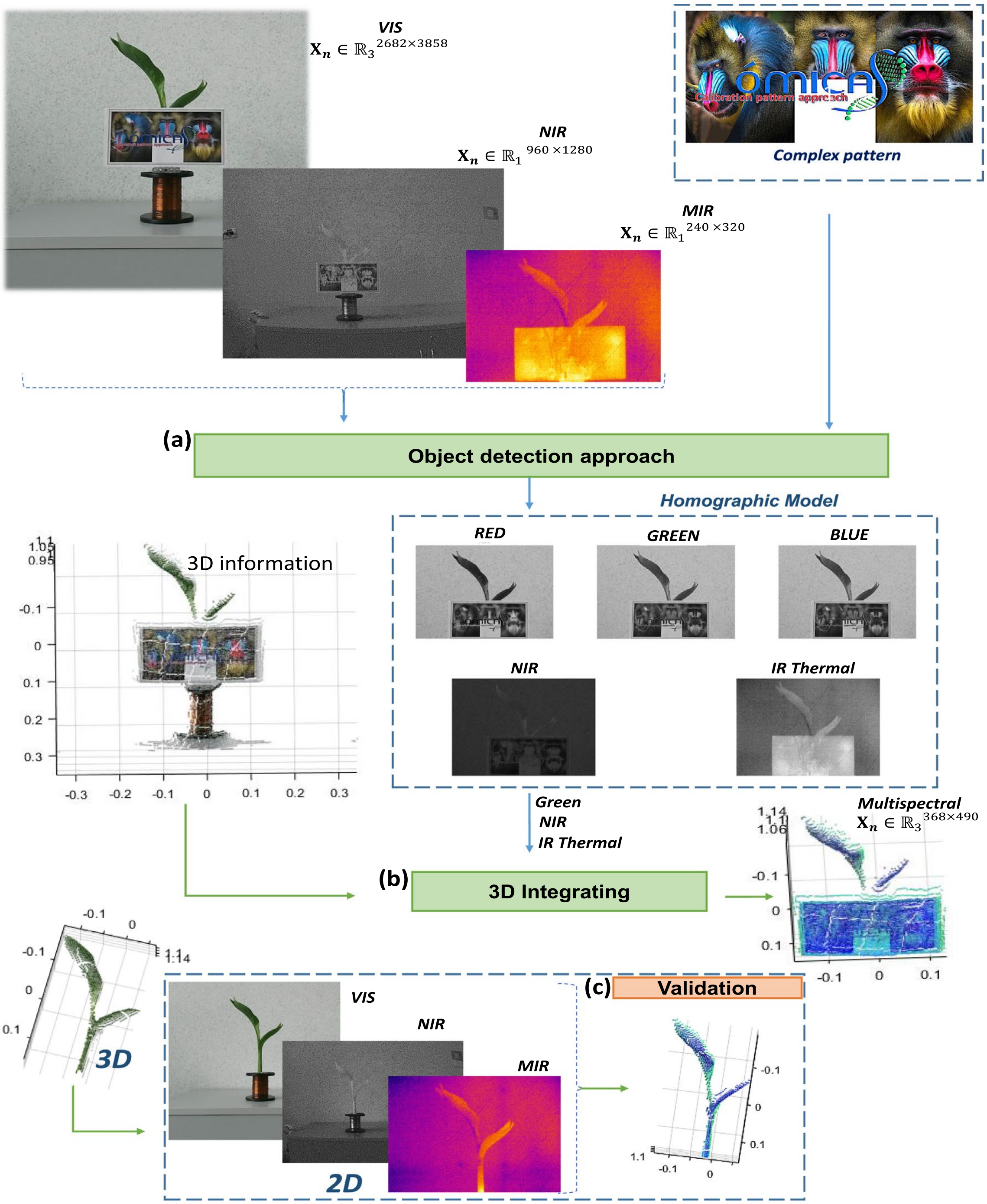

VIS
$\mathbf{X}_n \in \mathbb{R}_3^{2682\times3858}$
NIR
$\mathbf{X}_n \in \mathbb{R}_1^{960\times1280}$
MIR
$\mathbf{X}_n \in \mathbb{R}_1^{240\times320}$
OMICAS
Calibration pattern approach
Complex pattern
(a)
Object detection approach
Homographic Model
RED
GREEN
BLUE
NIR
IR Thermal
3D information
Green
NIR
IR Thermal
(b)
3D Integrating
Multispectral
$\mathbf{X}_n \in \mathbb{R}_3^{368\times490}$
3D
VIS
NIR
MIR
2D
(c)
Validation

Fig. 19 Sensory Fusion, going from 2D to 3D. (**a**) Object detection approach (**b**) 3D integrating (**c**) Validation

Results and Discussion

The validation of the proposed methodology involves reproducing matching results and integrating information from all sensors into the 3D reference frame. This results in a three-dimensional model composed of data from three cameras, representing a multi-sensory fusion approach over multispectral wavelengths.

In Fig. 19a, images captured in the laboratory with different sensors are displayed. These include three channels: the visible RGB spectrum (VIS), the multispectral near-infrared channel (NIR), and the thermal middle-infrared channel (MIR).

Figure 20 illustrates the resulting channels within the 3D reference frame. Further, the 3D information is depicted in Fig. 21a, b, integrating spectral bands from the various cameras: (i) a channel from the visible RGB spectrum of the 3D sensor, (ii) the multispectral NIR channel of the infrared camera, and (iii) the infrared channel of the thermal camera.

In Fig. 22, the homographic transformation relating the infrared image to the RGB image is depicted. The resulting image combines common regions, as shown in Fig. 23. This clipping represents a new channel available for generating an integrated image along with the RGB image.

Finally, Fig. 24 presents an image composed of the RG channels of the Kinect sensor (depicted in the red box in Fig. 22), along with (a) the NIR channel of the Parrot sensor

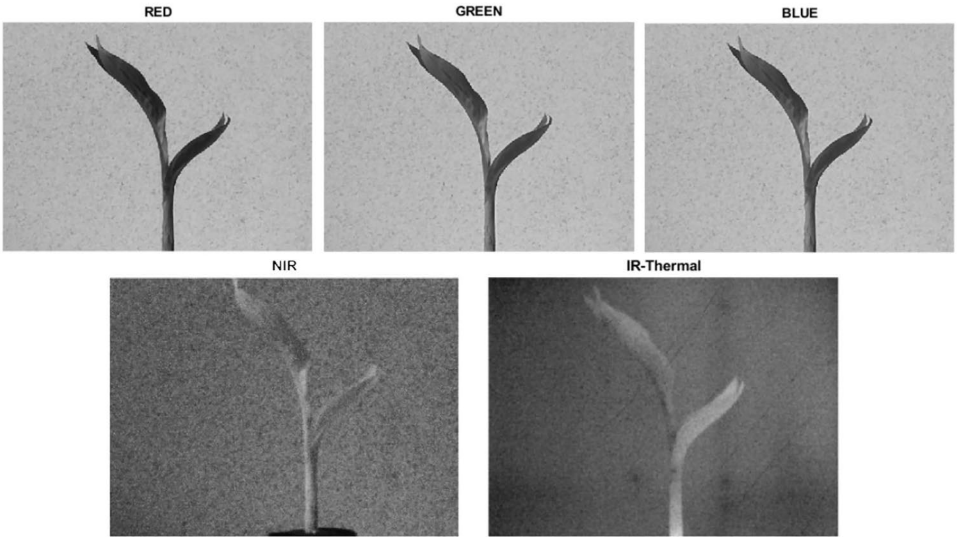


Fig. 20 Five channels in the same frame, RGB from the 3D camera, NIR from the multispectral camera, and IR-Thermal from the thermal camera. Resolution = [368×490]

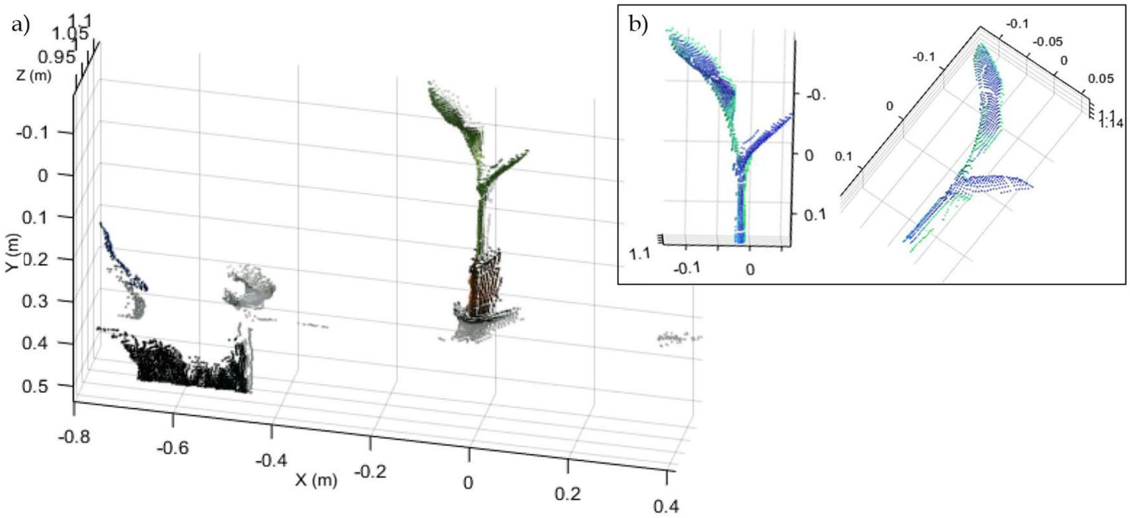


Fig. 21 **a** RGB initial 3D model. Captured with the Kinect v02 sensor. **b** Resulting 3D model, composed of: (i) the G channel of the 3D sensor, (ii) the IR channel of the parrot camera and (iii) the infrared thermal channel of the Fluke thermal camera

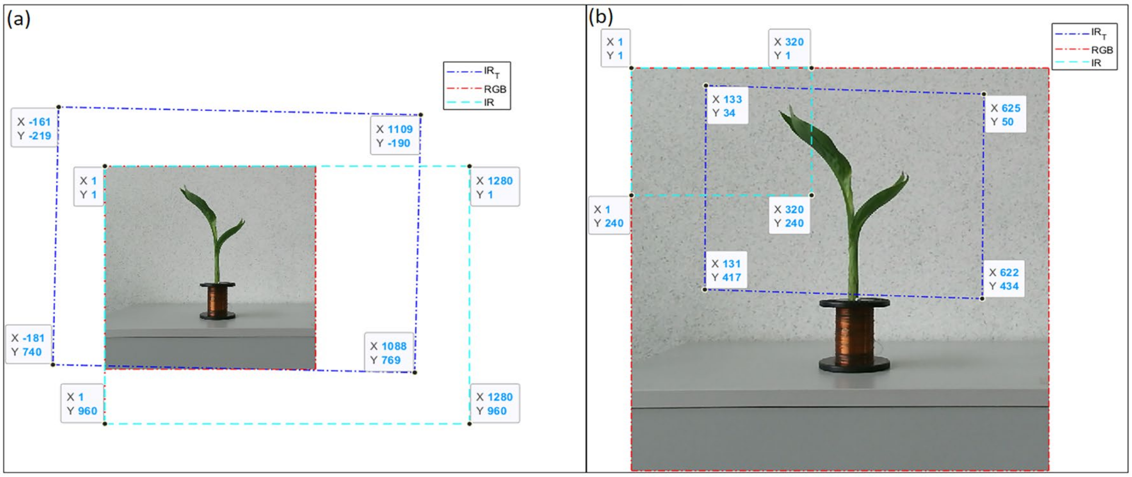


Fig. 22 Homographic projection between the Kinect camera and: (**a**) the NIR multispectral camera and (**b**) the NIR thermal multispectral camera

Fig. 23 Homographic transformation of the: (**a**) NIR multispectral image and (**b**) MIR multispectral image

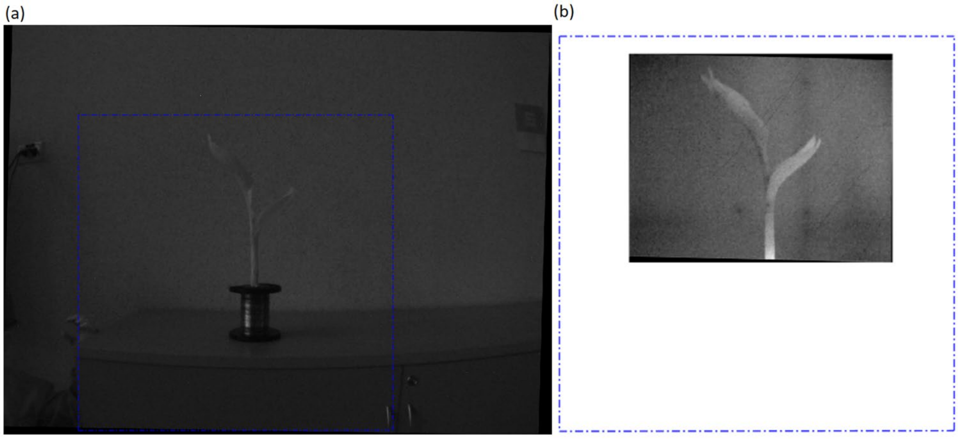
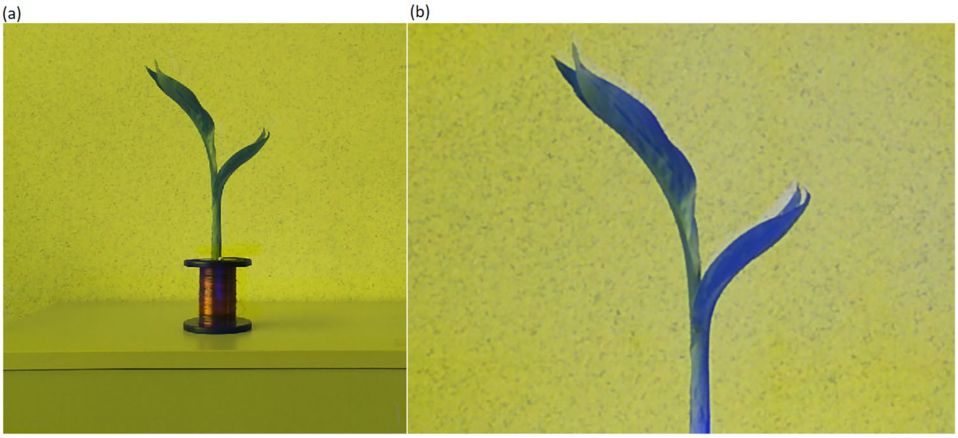


Fig. 24 (**a**) RGN Image with Resolution = [757×740]. (**b**) RGT Image with Resolution = [368×490]



and (b) the MIR channel of the Fluke sensor (shown in the blue box in Fig. 23).

The images integrating multispectral information in Fig. 20, for instance, such as the one depicted in Fig. 24, which combines VIS and NIR-MIR information, are displayed in pseudo color in the RGB standard. This coloring is necessary because the human visual system cannot perceive these wavelengths. In Fig. 24, we observe the result of the investigation: an image composed of the green visible channel (G) along with two multispectral channels, MIR and NIR, referenced in 3D space.

The resulting image size varies based on the common regions captured by the cameras in the scene. The sensory fusion strategy, employing pattern recognition and statistical optimization, effectively models the capture of different images to generate a single, multisensory integrated image. This approach, crucial for plant-related applications, performs object recognition once during calibration and is capable of overcoming occlusions. Furthermore, it offers the potential to propose new robust descriptors.

The success of the proposed sensory fusion, as depicted in Fig. 19, is evident. Initially, the projective model implements a complex pattern recognition approach, followed by validation using multi-camera scene acquisition. The result is a point cloud composed of multispectral information initially in 2D, demonstrating the efficacy of the strategy.

The effectiveness of the proposed sensory fusion, as depicted in Fig. 19, is appreciable. It begins with the implementation of a complex pattern recognition approach in the projective model, followed by validation through multi-camera scene acquisition. This results in a point cloud initially containing 2D multispectral information, demonstrating the strategy's success.

Conclusions

This paper proposes a novel multi-camera sensory fusion technique for complex object detection based on homography model transformations. The proposed technique uses probabilistic optimization and pattern recognition methods that are widely used in pose estimation, odometry, and tracking tasks. The technique performs sensory fusion in a multisensory setting, which is a novel contribution.

The technique produces homographic transformations for each camera, using images with rich patterns that do not require a large-scale dataset. This benefit allows for detailed morphological modeling of individual plants or crops, as the cameras are close to the target, where comprehensive datasets may not be available. The technique performs object recognition only once during the initial setup, and then applies sensory fusion to unknown scenes. The technique generates a 3D representation from initially 2D multispectral images, which include near or thermal infrared information. However, the technique is sensitive to distance changes, and needs re-calibration if altered.

Future work will be oriented towards implementing and comparing convolutional neural networks for object recognition to enhance the sensory fusion performance. This will involve integrating homographic transformations, with a main challenge being the creation of a large image database.

The resulting model combines 3D information, facilitating precise morphological plant measurements, and multispectral data, enabling assessment of plant conditions such as weather stress or nitrogen deficiency. The sensory fusion approach generates valuable information for crop modeling applications, with the potential to extract morphological variables such as number and size of leaves, stems or plant height, in addition multispectral information at the plant scale, it is non-invasive that can be installed in agricultural production fields.

Funding Open Access funding provided by Colombia Consortium. This work was partly funded by Pontificia Universidad Javeriana in Bogota Colombia, under the project ID 20366 -Optimized navigati on control applied to monitoring and inspection. Also, by the Omics Science Research Insti tute (iOMICAS) anchored in Pontificia Universidad Javeriana in Cali Colombia.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. Copyright. In: Bhullar, G.S., Bhullar, N.K. (eds.) Agricultural Sustainability, p. Academic Press, San Diego (2013). https://doi.org/10.1016/B978-0-12-404560-6.00017-4. https://www.sciencedirect.com/science/article/pii/B9780124045606000174
2. Rose MT, Rose TJ, Pariasca-Tanaka J, Widodo, Wissuwa M. Revisiting the role of organic acids in the bicarbonate tolerance of zinc-efficient rice genotypes. Funct Plant Biol. 2011;38(6):493–504. https://doi.org/10.1071/FP11008. (**cited By 32**).
3. Wu C, Zou Q, Xue S, Mo J, Pan W, Lou L, Wong MH. Effects of silicon (si) on arsenic (as) accumulation and speciation in rice (oryza sativa l.) genotypes with different radial oxygen loss (rol). Chemosphere. 2015;138:447–53. https://doi.org/10.1016/j.chemosphere.2015.06.081.
4. Wu C, Zou Q, Xue S-G, Pan W-S, Yue X, Hartley W, Huang L, Mo J-Y. Effect of silicate on arsenic fractionation in soils and its accumulation in rice plants. Chemosphere. 2016;165:478–86. https://doi.org/10.1016/j.chemosphere.2016.09.061.
5. Zhang L, Yang Q, Wang S, Li W, Jiang S, Liu Y. Influence of silicon treatment on antimony uptake and translocation in rice genotypes with different radial oxygen loss. Ecotoxicol Environ

Saf. 2017;144:572–7. https://doi.org/10.1016/j.ecoenv.2017.06.076.

6. Matsubara K, Yonemaru J-I, Kobayashi N, Ishii T, Yamamoto E, Mizobuchi R, Tsunematsu H, Yamamoto T, Kato H, Yano M. A follow-up study for biomass yield qtls in rice. PLoS ONE, 2018;13(10). https://doi.org/10.1371/journal.pone.0206054. cited By 2
7. McCouch WMTCS. Open access resources for genome-wide association mapping in rice. Nat Commun. 2016;7:1. https://doi.org/10.1038/ncomms10532.
8. Bouman BAM, Peng S, Castañeda AR, Visperas RM. Yield and water use of irrigated tropical aerobic rice systems. Agric Water Manag. 2005;74(2):87–105. https://doi.org/10.1016/j.agwat.2004.11.007.
9. Kamffer KAOA Z Bindon. Optimization of a method for the extraction and quantification of carotenoids and chlorophylls during ripening in grape berries (vitis vinifera cv. merlot). Journal of Agricultural and Food Chemistry, 2020;58. https://doi.org/10.1021/jf1004308
10. Ling Q, Wang S, Ding Y, Li G. Re-evaluation of using the color difference between the top 3rd leaf and the 4th leaf as a unified indicator for high-yielding rice. Sci Agric Sin. 2017;50(24):4705–13. https://doi.org/10.3864/j.issn.0578-1752.2017.24.004. (**cited By 2**).
11. Colorado JD, Calderon F, Mendez D, Petro E, Rojas JP, Correa ES, Mondragon IF, Rebolledo MC, Jaramillo-Botero A. A novel nir-image segmentation method for the precise estimation of above-ground biomass in rice crops. PLoS ONE. 2020;15(10):6.
12. Correa ES, Calderon F, Colorado JD. Gfkuts: A novel multispectral image segmentation method applied to precision agriculture. In: 2020 Virtual Symposium in Plant Omics Sciences, OMICAS 2020 - Conference Proceedings, 2020. Cited By :2
13. Jing Z, Guan H, Zhao P, Li D, Yu Y, Zang Y, Wang H, Li J. Multispectral lidar point cloud classification using se-pointnet++. Remote Sens. 2021;13(13):8.
14. Jimenez-Sierra DA, Correa ES, Benítez-Restrepo HD, Calderon FC, Mondragon IF, Colorado JD. Novel feature-extraction methods for the estimation of above-ground biomass in rice crops. Sensors. 2021;21(13):4369.
15. Yang J, Song S, Du L, Shi S, Gong W, Sun J, Chen B. Analyzing the effect of fluorescence characteristics on leaf nitrogen concentration estimation. Remote Sens. 2018;10:9. https://doi.org/10.3390/rs10091402.
16. Yuan Z, Ata-Ul-Karim ST, Cao Q, Lu Z, Cao W, Zhu Y, Liu X. Indicators for diagnosing nitrogen status of rice based on chlorophyll meter readings. Field Crops Res. 2016;185:12–20. https://doi.org/10.1016/j.fcr.2015.10.003.
17. Yamane K, Kawasaki M, Taniguchi M, Miyake H. Correlation between chloroplast ultrastructure and chlorophyll fluorescence characteristics in the leaves of rice (oryza sativa l.) grown under salinity. Plant Prod Sci. 2008;11(1):139–45. https://doi.org/10.1626/pps.11.139.
18. Zhang H, Zhu L-f, Hu H, Zheng K-f, Jin Q-y. Monitoring leaf chlorophyll fluorescence with spectral reflectance in rice (oryza sativa l.). Proc Eng. 2011;15:4403–8. https://doi.org/10.1016/j.proeng.2011.08.827. (**CEIS 2011**).
19. Subhash N, Mohanan CN. Laser-induced red chlorophyll fluorescence signatures as nutrient stress indicator in rice plants. Remote Sens Environ. 1994;47(1):45–50. https://doi.org/10.1016/0034-4257(94)90126-0. (**Fluorescence Measurements of Vegetation**).
20. Liu S. Phenotyping wheat by combining adel-wheat 4d structure model with proximal remote sensing measurements along the growth cycle. PhD thesis, 2016.
21. Polder G, Hofstee JW. Phenotyping large tomato plants in the greenhouse using a 3D light-field camera, vol. 1, pp. 153–159. American Society of Agricultural and Biological Engineers, ???, 2014.
22. Sandhya Devi RS, Vijay Kumar VR, Sivakumar P. A review of image classification and object detection on machine learning and deep learning techniques. In: Proceedings of the 5th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2021, 2021.
23. Qingyun F, Zhaokui W. Cross-modality attentive feature fusion for object detection in multispectral remote sensing imagery. Pattern Recogn. 2022;1:30.
24. Lin T-, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common Objects in Context. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8693 LNCS, pp. 740–755 (2014). Cited By :11409
25. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int J Comput Vis. 2010;88(2):303–38 (**Cited By :8991**).
26. Gani MO, Kuiry S, Das A, Nasipuri M, Das N. Multispectral Object Detection with Deep Learning. Communications in Computer and Information Science, vol. 1406 CCIS, pp. 105–117 (2021). Cited By :3
27. Münzinger M, Prechtel N, Behnisch M. Mapping the urban forest in detail: From lidar point clouds to 3d tree models. Urban For Urban Green. 2022;74:2.
28. Li H, Zech J, Ludwig C, Fendrich S, Shapiro A, Schultz M, Zipf A. Automatic mapping of national surface water with openstreetmap and sentinel-2 msi data using deep learning. Int J Appl Earth Observ Geoinform. 2021;104:2.
29. Jurado JM, López A, Pádua L, Sousa JJ. Remote sensing image fusion on 3d scenarios: a review of applications for agriculture and forestry. Int J Appl Earth Observ Geoinform. 2022;11:2.
30. Wichmann V, Bremer M, Lindenberger J, Rutzinger M, Georges C, Petrini-Monteferri F. Evaluating the potential of multispectral airborne lidar for topographic mapping and land cover classification. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W5, 113–119 (2015). https://doi.org/10.5194/isprsannals-II-3-W5-113-2015
31. Kaygusuz N, Mendez O, Bowden R. Multi-camera sensor fusion for visual odometry using deep uncertainty estimation. 2021. https://doi.org/10.1109/itsc48978.2021.9565079.
32. Dockstader SL, Tekalp AM. Multiple camera fusion for multi-object tracking. In: Proceedings 2001 IEEE Workshop on Multi-Object Tracking, 2001;95–102. https://doi.org/10.1109/MOT.2001.937987
33. Cachique SM, Correa ES, Rodriguez-Garavito C. Intelligent digital tutor to assemble puzzles based on artificial intelligence techniques. In: International Conference on Applied Informatics, 2020;56–71 . Springer
34. Alam MS, Morshidi MA, Gunawan TS, Olanrewaju RF, Arifin F. Pose estimation algorithm for mobile augmented reality based on inertial sensor fusion. Int J Electr Comput Eng. 2022;12(4):3620–31.
35. Yang L, Li Y, Li X, Meng Z, Luo H. Efficient plane extraction using normal estimation and ransac from 3d point cloud. Computer Standards and Interfaces, 2022;82. Cited By :1
36. Gao L, Zhao Y, Han J, Liu H. Research on multi-view 3d reconstruction technology based on sfm. Sensors. 2022;22:12.
37. Correa ES, Parra CA, Vizcaya PR, Calderon FC, Colorado JD. Complex object detection using light-field plenoptic camera **1576 CCIS**, 2022;119–133.
38. Zhang C. Decoding and calibration method on focused plenoptic camera. Comput Vis Med. 2016;2:2096–662. https://doi.org/10.1007/s41095-016-0040-x.

39. O'brien S, Trumpf J, Ila V, Mahony R. Calibrating light-field cameras using plenoptic disc features. In: 2018 International Conference on 3D Vision (3DV), 2018;286–294. https://doi.org/10.1109/3DV.2018.00041
40. Edgar S Correa1, PRVFC Carlos A Parra1, Colorado JD. Complex object detection using light-field plenoptic camera, 2022;21:977–1000. https://doi.org/10.1016/S0262-8856(03)00137-9
41. Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 2004;60:91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.
42. Fotouhi HHK-NMAKS M. Sc-ransac: spatial consistency on ransac. Multimedia Tools and Applications, 2019;78(7):9429–9461. https://doi.org/10.1007/s11042-018-6475-6
43. Solem JE. Programming computer vision with python: Tools and algorithms for analyzing images, 2012. Pages 72-74. " O'Reilly Media, Inc."
44. Zhuang L, Yu J, Song Y. Panoramic image mosaic method based on image segmentation and improved sift algorithm, 2021;2113. Chap. 1
45. Luo X, Chen W, Du X. A matching algorithm based on the topological structure of feature points 2021;11720. Cited By :1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.