

Автономная некоммерческая образовательная организация
высшего образования «Сколковский институт науки и
технологий»

На правах рукописи

Осокин Илья Витальевич

**Робот и система компьютерного зрения для поиска
заболеваний и оценки объема урожая при мониторинге
теплиц**

**Специальность: 2.3.1. Системный анализ, управление
и обработка информации, статистика**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва — 2025

Работа выполнена в Автономной некоммерческой образовательной организации высшего образования «Сколковский институт науки и технологий».

Научный руководитель: **Осиненко Павел Валерьевич**,
доктор физико-математических
наук, доцент

Защита состоится **DD месяц ГОД г. в XX часов XX минут** на заседании диссертационного совета **Х.Х.Х.Х.**, созданного на базе Автономной некоммерческой образовательной организации высшего образования «Сколковский институт науки и технологий» (Сколтех)

по адресу: Территория Инновационного Центра «Сколково»,
Большой бульвар д.30, стр.1, Москва 121205

С диссертацией можно ознакомиться в библиотеке Сколтеха и на сайте организации <https://dissovet.skoltech.ru/>

Автореферат разослан «_____» _____ **2025** года.

**Ученый секретарь
диссертационного
совета**

к.ф.-м.н.

Копелевич Григорий Александрович

Autonomous Non-Profit Organization for Higher Education
“Skolkovo Institute of Science and Technology”

As a manuscript

Osokin Ilya Vital'evich

**A robot and a computer vision system for disease
detection and crop yield estimation in greenhouse
monitoring**

**Speciality: 2.3.1. System Analysis and Information
Processing**

**DISSERTATION ABSTRACT
of the dissertation for the Degree of
Doctor of Philosophy in **Engineering****

Moscow — **2025**

The work has been performed at the Autonomous Non-Profit Organization for Higher Education “Skolkovo Institute of Science and Technology”.

Scientific supervisor: **Osinenko Pavel Valer’evich,**
Doctor of Science in Physics and Mathematics, Associate Professor

The defense will take place on **DD Month Year at XX:XX p.m.** at the meeting of the Dissertation Council **X.X.X.X.**, based at Skolkovo Institute of Science and Technology (Skoltech)

address: Skolkovo Institute of Science and Technology, the territory of the Innovation Center “Skolkovo”, Bolshoy Boulevard, 30, bld.1, Moscow 121205, Russia

The text of the dissertation is available at the Skoltech library or on the website <https://dissovet.skoltech.ru/>

Dissertation Abstract was sent on «_____»_____2025 .

**Academic secretary of the
Dissertation Council**

Candidate of Physical and
Mathematical Sciences

Kopelevich Grigory Alexandrovich

General characteristics of the work

0.1 Relevance of the work

Modern farming is characterized by both scale and efficiency. Both precise yield estimation and disease detection are its integral parts. Tangerines and tomatoes are one of the most important fruit in the agricultural produce worldwide [1].

More specifically, knowing the mass of the tomatoes that are ready to be picked is crucial for proper logistics planning [2]. And volume is one of the most important physical attributes of the fruit yield in agricultural production. If performed manually, it is time-consuming, expensive and prone to error due to the loss of focus under the monotonous conditions of such labour. In order to automate this process, computer vision-based systems are often used. Because of the scale of modern farms and greenhouses, covering them entirely with stationary cameras is infeasible. Thus, the majority of the solutions are based on autonomous robots, i.e. flying or wheeled.

This work addresses the problem of estimating position, size and orientation of tomatoes and tangerines under the conditions close to the real greenhouse. This problem is addressed with two different approaches, one relying on classical algorithmical surface fitting, and another one being an end-to-end Neural Network-based approach.

Secondly, disease detection is crucial in loss prevention and taking timely measures to optimize the production. There are multiple diseases that manifest in a visible way, with powdery mildew being one of them. In this work the the problem of powdery mildew identification is addressed. In order to solve it, a robot was developed, as well as a multi-camera setup for data collection.

0.2 Dissertation Goals

The goal of this dissertation is to investigate the performance of volume estimation algorithms that process 3D point cloud data in application to ellipsoidal objects in agricultural setting. The second goal it to evaluate the performance of the disease identification algorithms in application to the tomato plants.

In order to achieve these goals, the following problems are solved.

Thie first is to develop an implementation of the ellipsoid fitting algorithm with point cloud input. The second is to evaluate its performance in multiple settings, including various data collection conditions and different objects. The third is to solve the problem of volume estimation with an end-to-end approach, i.e. Neural Network. The fourth is to perform data collection in real greenhouse conditions, as well as markup, data coherency evaluation and model training for the problem of disease detection. The fifth is to develop a robot capable of moving in the greenhouse on two types of surface, while carrying a computer and a camera setup, with sufficiently long battery lifetime.

0.3 Scientific Novelty

The following main results should be highlighted.

- A problem of tomato and tangerine volume estimation was solved on the real point cloud data.
- A problem of powdery mildew identification on tomato leaves was solved on real RGB data from user-grade cameras
- PointNet was adapted to the regression problem and used to estimate the volume of tangerines in the end-to-end fashion

0.4 Practical Outcomes

The robot that was developed during the research was certified to be TRL-5 (Technology Readiness Level) in terms of its ability to move on both flat surface and rails. The robot is also capable of using its kinematic model to follow any curve on a two-dimensional surface. Practically it means that the platform is ready to carry scientific or industrial load, if controlled in the environment properly.

Three patents were issued to Skoltech during the research:

1. A dataset of tomato leaves infected by powdery mildew.
The unique feature is that the images were filmed with a mid-price RGB camera from a moving robot in the real greenhouse, which is in great contrast with the publicly available datasets, filmed in the lab, with artificial lighting, steady scene and expensive camera setups
2. A computer program capable of classifying images into infected by powdery mildew and not infected
3. A wheel, corresponding of a Mecanum wheel, rail wheel, mounted coaxially, integrated motor, and a suspension system

Finally, a company AIDA Robotics is in process of being established.

0.5 Research methodology

The methodology incorporates theoretical ideas from the field of pattern recognition. The design of the main algorithms is based on random sample consensus-based approaches, as well as the properties of the second-order curves and surfaces. At the same time, a lot of the specific design solutions both in terms of the software and the hardware were made in accordance with the real-world demands from the engineering standpoint. It provided the means to consider the prospective usage in terms of vibration reliability, autonomous operation time, and locomotion abilities of the robot.

The following is a list of main software tools and frameworks used to conduct numerical experiments:

- OpenCV
- Open3D
- ROS2
- Matlab
- PyTorch

0.6 Propositions for Defense

1. The problem of model evaluation for the second order surface fitting was reduced to a matrix multiplication. The experimental results suggest that the performance of the proposed method on a single modern laptop allows for the real-time greenhouse monitoring.
2. The influence of the noise on the quality of ellipsoid fitting was numerically evaluated. The practical outcomes include results for the second-order curves as a simpler object.
3. A unique dataset of tomato leaves infected by powdery mildew was collected, marked and patented. The defining feature is that the images were taken with a user-grade camera from a moving robot in the real greenhouse. The publicly available datasets are all filmed in lab conditions with artificial lighting, steady objects and expensive cameras.
4. A robot was developed and its crucial feature (the wheel) was patented. The robot was specifically designed to serve as a hardware support for the whole project, and it encompasses the camera setup, the computing capabilities onboard, and a prolonged lifetime on a single charge.

0.7 Validity and approbation

The results were presented on 4 conferences: MIPT conference, Zavalishinsky readings, and two papers at EDM conference.

0.8 Personal contribution of the author

The author actively contributed to several research activities.

First, the autor formulated the research hypothesis, proposed and validated experimental setups, as well as participated in experiments, including ones in the real greenhouse environment.

Second, the author participated in the algorithms development and numerical experiments.

Third, the author took part in the designing, manufacturing, and overall working on the omnidirectional robot, including milling, assembly, painting, components purchase, cable management, soldering.

Fourth, the author participated in writing all 6 papers, that are the foundation of the thesis presented.

Fifth, the author made significant contribution to the legal part of the project, in particular grant-related activities, reports, patenting, and TRL.

Sixth, the author contributed to the organization of the project, including personnel search, trips, and procurement.

Dissertation Summary

Modern agriculture is characterized by scale and efficiency. Autonomous disease detection as well as autonomous yield estimation are crucial to the automatization, loss prevention and further efficiency improvement. As some of these labour-intensive tasks are being automated, certain technical problems should be solved. In this work a complex solution is presented, including a robot with omnidirectional shassis and a camera setup, and a set of algorithms that solve the problems of powdery mildew identification and yield volume estimation for ellipsoidal fruit and vegetables. An improvement was proposed to the algorithm of ellipsoidal model fitting to the point cloud data. Specifically, the process of checking the hypothesis was sped up. All the proposed algorithms were extensively

tested on real-world data and evaluated in terms of the metrics relevant to the industry. The results suggest the applicability of the proposed approaches in the real-world scenarios.

In the recent years computer vision-based monitoring methods gain traction. They mainly rely on RGB sensors, and in certain cases depth cameras. While RGB-only sensors are much cheaper, depth information about the scene allows for more precise volume estimation. One of the key factors in the vision-based volume estimation is the noise that is an inherent property of the sensor data in the real greenhouse facility. In order to accommodate for these noises, robust volume estimation techniques should be applied.

Let us describe the classical RANSAC algorithm, as per paper by Fischler and Bolles. It was previously mentioned, that least squares suffer from sensitivity to outliers, and Hough transform from the rapid growth of the accumulator size. This is where random sample consensus-based algorithms could be used. In contrast to the Hough transform, where all the parameter space is considered explicitly, random sample consensus relies on the chosen models from all the possible ones from the parameter space, thus saving memory. At the same time, in contrast to the least squares methods, random sample consensus is not obliged to take into account all the points of the input data. It relies on the sampling of a number of subset of data. For each subset, a separate model is created. The number of the points in this subset has to be such that it determines a single unique model for lines. It will be two points for a line, and for ellipse it will be five points. After the model is obtained, it is calculated how many of the input data points are described by the model.

This notion of how the data points are described by the model is in itself a variety of different approaches. The first way to tell if the data point is described by a model is to measure the distance from this point to this model, and the second way of doing this relies on the so-called algebra distance. Considering a case of an ellipse, the equation describing it is a second-order equation with two independent variables. If the left side of this equation is considered as a polynomial, the value of this polynomial is zero on the ellipse

and not zero inside and outside. The value of the polynomial on the data point could be considered as a distance from the point to the model. There are also alternative approaches to the measurement of the distance from the point to the model. In particular, there is a mixture of the geometrical distance, and the distance measured along the semiaxes of the ellipsoid in the paper.

There are also ways to count the fitness of the model for the dataset. After the distances are obtained with one of the methods described above, a single scalar metric should be calculated. It could be a sum of indicator functions. If the distance from the point to the model is below a certain threshold, the point is considered to be fit for the model. The best model among the same ones is the one that has the biggest number of inliers. However, there are alternative approaches to counting the fitness of the model for the data. For instance, monotonous decreasing function could be used as a weight function instead of the threshold function, such as the exponential of minus distance from the point to the model.

A method of ellipsoid volume estimation by noisy point cloud data is proposed. It relies on the combination of RGB and depth data. Specifically, it consists of two parts: detecting the objects in the RGB image, that is registered to the point cloud from a depth camera, and fitting ellipsoids to the corresponding subsets of that point cloud. The proposed method was tested on synthetic data and real data in agricultural context. Its main feature is that it is suitable for the real-world environments. For industrial applications, such as modern tomato greenhouses, it means that the tomatoes are not supposed to be manually picked. In contrast to many modern approaches, the proposed one does not rely on the contrastive background or an expensive setup with multispectral cameras. For the real experiments Intel RealSense D435i camera was used. The method’s performance was evaluated on both synthetic and real data in terms of the Intersection over Union (IoU) between the predicted ellipsoid and the real one, volume error and processing time. The dependence of the output quality on the parameters, such as number of RANSAC iterations and inlier threshold, is presented.

The method is compared with three other modern algorithms, showing superior performance on the complex real-world data. It is lightweight enough to be used on an autonomous agricultural robot with a user-grade laptop on board.

Modern agriculture poses several challenges in terms of the requirements for monitoring speed and precision. Certain greenhouses of the major agricultural companies are so huge, that it is virtually impossible to timely examine the whole facility with a reasonable human workforce. While it is possible to introduce more and more human workers, a sustainable long-term solution is to use autonomous robots to automate such repetitive tasks.

While robot-assisted agriculture has been developing for a long time already, the integration of a robot into an industrial-scale production often demands substantial effort. It requires the robot to have a high degree of autonomy while keeping the modifications that the greenhouse facility should go through to the minimum.

The target greenhouses for this project are rectangular 120000 m² (12 hectares) buildings with glass rooftops. During the night and in cloudy weather red-blue LEDs are used to provide sufficient lighting. The temperature and humidity are kept stable at all times. Several types of tomatoes are grown there with the help of hydroponics. One tomato plant grows up to 35 meters in length during 10 months of cultivation. The majority of the length of the plant is kept horizontal, and the top of the plant is at a height of approximately 4 meters.

The density of the biomass in these greenhouses allows for cost savings because of the scale of the production. On the other hand, it leads to the rapid spread of various kinds of bacterial infections, fungi, viruses, and insects. Thus, timely monitoring in such greenhouses is crucial. Powdery mildew is known to infect the lower leaves first and then spread up the plant. It develops best at temperatures slightly below 30°C and high humidity, which are precisely the conditions that the tomatoes are grown in for the great portion of the production cycle. Thus, the development of the powdery mildew is rapid, taking 4 to 7 days to progress from early to medium stage.

With the aforementioned requirements for monitoring speed in mind, an autonomous robot on an omnidirectional platform was developed and manufactured. It is equipped with cameras, covering the whole height of the tomato plant, and an onboard computer to process the data in real-time.

This paper presents a work on the data acquisition with this robot, as well as coordinating the markup procedure by human experts and a comparison of the performance of several neural networks.

For the problem of tomato greenhouse monitoring the main requirements for the vision system are the following.

First, the combination of the Fields of View (FoV) of the cameras should cover the entire height of the plants, from the tomatoes to the tips of the vegetation. Diseases, parasites, and damage can occur anywhere on the plant. This requirement is addressed by installing a number of cameras, their proper positioning, and their parameters. Three web cameras were utilized. The sketch of this setup is given in Figure, and the real assembly is given in Figure. At this point, only one side is covered by combined FoVs, but it is planned to replicate that and cover the other side as well.

Second, the image processing should be performed in real-time. Ideally, the robot is expected to spend all the time in the field moving. With real-time processing, it is possible for the workers to rapidly examine specific rows of tomato plants. The robot should be capable of observing the same part of the plant from different locations, meaning that the data processing should happen in real-time with a certain overlap in the frames.

During the design of the vision subsystem, the following assumptions were adopted. In the initial testing, the robot was moving at a speed of 0.26 m/s, which is a safe value for indoor applications. Considering potential acceleration for the sake of faster monitoring, an upper-speed limit could be set to 0.5 m/s. The horizontal FoV for the chosen optics covers 1.07 m. Thus, moving an object through it from left to right takes nearly 2 s. To assure the 5-fold overlap, each of the cameras has to capture an image at

least once every 0.4 s. That results in the requirement for the processing speed of 8 frames of 1056×1056 pixels per second. This requirement is met with a huge margin.

Overall, a work was done in the field of automated greenhouse monitoring. The main contribution lies in the intersection of theoretically supported developments in the surface recognition and engineering problems that were solved during this work.



Figure 1 — The third version of the robot with the camera masts installed. The height of the robot is approximately 3.5 meters, allowing it to inspect the whole tomato plant in one go.

The main results of this dissertation are:

1. A method of ellipsoid volume estimation was developed and tested in the real environment. Synthetic data was generated with gradual increase in complexity. A real dataset

of point clouds and corresponding RGB images was gathered and marked. The dependence of the quality of the method’s outout on the iterations number was evaluated in numerical experiments, showing minimal average relative volume estimation error of 0.25.

The comparison of the method with three alternative approaches was made, showing superior performance on the real data both in terms of IoU and volume error. It was thus shown that the proposed method can be applied to the real-world ellipsoid volume estimation.

2. A method of robust volume estimation was applied to tangerines. A dataset of point clouds and RGB images was gathered, and the tangerines were measured in terms of mass and volume. Numerical experiments were conducted with varying hyperparameters. The results suggest that Random Sample Consensus can be successfully applied in the problem of volume estimation.

The proposed approach has a number of applications in agriculture. First, it can be applied to the problem of the instant yield estimation. In order to assess the volume of the yield, a robot can be used that will monitor the agricultural facility and provide the estimated volume. Second, the measurements of the volume across different time periods can be helpful for the development of a prognostic tool for the prospective yield.

Future work can include the generalization of the method to other types of fruit, including non-elliptical ones. In particular, they can have a shape of a curved ellipsoid, like a banana, or a superellipsoid, like sweet pepper.

The proposed method can run on a normal computer with resonably computational load. In future RANCAS for ellipsoids can be adapted to the CUDA-based inference in order to speed up the computations.

3. A neural network-based end-to-end tangerine volume estimation method was proposed. It relies on the PointNet++ architecture. The training on the custom dataset takes

nearly 1.5 hours, and the error in the volume estimation lightly exceeds 10%.

The inference time of 290 milliseconds allows for the real-time onboard inference with limited computational resources. The proposed approach can be used on mobile robots for the agricultural facilities monitoring.

4. An autonomous agricultural robot on an omnidirectional platform was designed, prototyped, and tested in a real environment. The camera's positioning allows for the examination of the whole tomato plant, and the onboard computer is capable of processing the data in real-time.

A dataset of powdery mildew-infected tomato leaves was collected, labeled and assessed in terms of consistency. It contains a sufficient amount of data for training modern neural networks of reasonable size, which makes it useful for further industrial applications.

Several modern neural networks were trained for classification and compared on two test sets with one of them being marked by a consensus of experts. The performance of the models is sufficient and matches with the mutual consistencies of the human experts.

It was demonstrated that real-time disease recognition could be performed on the robot while moving with the means of the user-grade cameras. The main limitation of the chosen approach is that RGB cameras require sufficient lighting to function properly. However, in the given circumstances this requirement is naturally met.

Overall, the work advances the ongoing endeavour of automating the greenhouse monitoring. There are still problems to be addressed, in particular they are connected to the integration of the proposed technologies into the on-site practice in the real greenhouses.

The next steps in this project are planned to be the following.

First, high-level control should be implemented in the system for the robot to be capable of localizing itself and autonomously following the trajectory set by the user.

Second, a Graphical User Interface should be implemented for the end user to control the robot without substantial UNIX knowledge.

Third, more disease types and crop types should be introduced into the vision subsystem.

References

1. MetaFruit meets foundation models: Leveraging a comprehensive multi-fruit dataset for advancing agricultural foundation models / J. Li [et al.] // Computers and Electronics in Agriculture. — 2025. — Vol. 231. — P. 109908.
2. *Tahir, G. A.* A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment / G. A. Tahir, C. K. Loo // Healthcare. Vol. 9. — MDPI. 2021. — P. 1676.

Author's publications on the dissertation topic

1. **Ilya Osokin**, Ryakin, Ilya and Moghimi, Sina and Patrikeev, Mikhail and Barsky, Ilya and Osinenko, Pavel, Neural network-based classification for automated powdery mildew detection in modern tomato greenhouses, IEEE Access, 2024, DOI;
2. **Ilya Osokin**, Ryakin, Ilya and Yaremenko, Grigory and Moghimi, Sina and Davidenko, Sergei and Guneavoy, Vladimir and Patrikeev, Mikhail and Osinenko, Pavel, The Impact of Noise on the Quality of the Second-Order Curve Recognition by Random Sample Consensus Algorithm, Proceedings of the Zavalishinsky Readings, 2025, DOI;
3. **Ilya Osokin**, Ryakin, Ilya and Moghimi, Sina and Davidenko, Sergei and Guneavoy, Vladimir and Yaremenko, Grigory and Osinenko, Pavel, Tangerine Volume Estimation by Point Cloud Data with Neural Networks, Proceedings of the EDM Conference, 2025, DOI;
4. Ilya Ryakin, **Ilya Osokin**, Moghimi, Sina and Davidenko, Sergei and Guneavoy, Vladimir and Yaremenko, Grigory and Osinenko, Pavel, Tangerine Volume Estimation by RANSAC on Point Cloud Data, Proceedings of the EDM Conference, 2025, DOI;
5. Ryakin, Ilya and Moghimi, Sina and Guneavoy, Vladimir and Davidenko, Sergei and Patrikeev, Mikhail and **Ilya Osokin** and Osinenko, Pavel, On the RANSAC performance for quadric surfaces in noisy conditions, IEEE Access, 2025, DOI;
6. Ilya Ryakin, **Ilya Osokin**, Osinenko, Pavel, Ellipsoidal objects volume estimation by combining detection with RANSAC, Proceedings of the MIPT Conference, 2025, DOI;