Implementation of Vision-Based Localization in Barelang-V Robot

Bernard Marcellino Sitio[1], Rifqi Amalya Fatekha[2] *

[1]Jurusan Teknik Elektro, Prodi Teknik Rekayasa Robotika, Politeknik Negeri Batam, Batam, Indonesia

*Email: saragihmarcel34@gmail.com

Received on dd-mm-yyyy | *Revised on dd-mm-yyyy* | *Accepted on dd-mm-yyyy*

***Abstract*—Reliable localization is critical needs for autonomous mobile robots in indoor and un even environment where GPS and wheel odometry are unreliable. This study implements vision-based localization system using RTAB-Map on the Barelang-V mobile robot, develop for the ABU Robot Contest. The system relies on an intel RealSense D455i RGB-D camera with built-in Inertial Measurement Unit (IMU). Visual-inertial data are processed within the ROS framework. Experimental results in an indoor environment demonstrate that the proposed approach achieves accurate and stable localization, with average positional errors of 0.051 meters when compared to ground truth measurements. These results indicate that the visual based localization with RTAB-Map pipeline provide effective solution for indoor mobile robot navigation.**

I. INTRODUCTION

The demand of robotics has been increasing significantly in the last few years, along side the decrease of human labor. To improve output, operational, effectiveness and security within the manufacturing settings, autonomous vehicles have become advantageous and effective solution. These automated units require reliable algorithm to perceive the surroundings and operate independently without human intervention [1].

Accurate and dependable localization is a fundamental requirement for autonomous mobile robot in the indoor environment[2]. The ability of a robot to estimate its position while navigating directly effects the performance of higher-level tasks such as path planning, obstacle avoidance and autonomous decision-making [3]. Traditional localization techniques make use of wheel odometry or ground contact to estimate movement [4]. But in some competition or specific environment like ABU robot contest arena robots may face wheel slipping, an uneven floor and operational rules that prohibit direct contact with the ground. These limitations limit the applications of traditional dead-reckoning-based methods and make alternatives such as position estimation independent of wheel feedback or line markers desire able[5].

A robot equipped with a vision sensor to estimate the position and orientation of the robot to perform SLAM called VSLAM [6]. There are quite a few algorithms that capable for estimating robot's pose using visual information, including Visual Odometry (VO) [7], Visual Inertial Odometry (VIO) [8][9], visual simultaneous localization and mapping (V-SLAM) [10]. Among various Visual SLAM frameworks,
RTAB-Map has demonstrated good performance in largescale and a long-term environment and also easier to implement [11][3]. Furthermore, RTAB-Map is well integrated with the Robot Operating System (ROS), which facilities deployment and integration to another robotic component such as microcontroller. RTAB-Map fuses visual features, depth information, and optional inertial data to build a global constant map while localizing the robot. This makes it appropriate for applications demanding autonomous navigation in constrained indoor environment where GPS is not accessible and wheel odometry is not trustworthy.

In this study, RTAB-Map is used in Barelang-V robot which was the competition robot from Barelang Robotic Team for ABU Robot Contest. The robot is equipped with an intel RealSense D455i depth camera system which combines a RGBD sensing and in built IMU. The depth and inertial measurement are fused together to enhance localization reliability, especially in fast robots' displacement or low-feature scenarios [12]. Without using wheel odometry or ground lines the system can make the robot design more flexible and have less mechanical dependence on ground observations by only depending on the D455i sensor.

Inspired by prior work on RTAB-Map localization and reconstruction [13], this paper presents the implementation of vision-based localization system on the Barelang-V mobile robot. The system relies on computation and intel RealSense D455i RGB-D camera for environmental perception. RTAB-Map is utilized to estimate the robot's pose in real time and provide localization feedback to the navigation module unlike approaches that focus on dense reconstruction or multi-sensor fusion, this work emphasizes a practical and deployable localization pipeline tailored for indoor mobile robots[14].

The main contribution of this research lies in the integration and evaluation of the RTAB-Map framework on the Barelang-V robot, highlighting its effectiveness for real-time indoor navigation using a single RGB-D sensor.

II. METHOD

The objective of this research is to enable a mobile robot's planar position (x, y) and heading orientation to support navigation. Achieving this requires a robust localization approach capable of operating in real-time and can operate in indoor environment, where the arena contain uneven terrain obstacle and relatively tall, which can limit the use of wheel

A. Overview

In this section is explained about the project system as a whole. System integration is achieved by connecting the Intel RealSense D455i camera to the processing unit (computer) using a single USB-C connection for data transmission.

The system uses an Intel RealSense D455i sensor because of depth measurement and integrated inertial sensing. The ideal range for the camera to get the right depth is about 3 meters, which is sufficient for getting visual and geometric features for localization using RTAB-Map. The combination of RGB data, depth information enables robust feature tracking and estimating motion, especially in indoor environment and ABU Robot contest arena. Representative RGB illustrated in Fig 1 and depth images captured by the sensor are illustrated in Fig 2.

Fig 1. D455i display physical scene

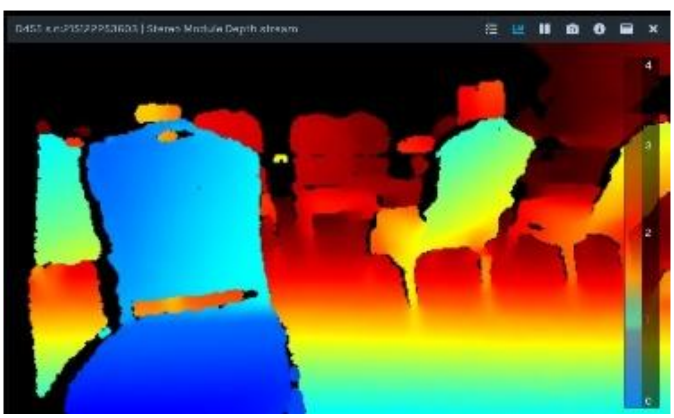

Fig 2. D455i display depth scene

The computational workload on this system is handled by a processing unit equipped with a quad-core processor with a base frequency of 2.50 GHz. To support multitasking and memory-intensive computer vision algorithms, the system uses 8 GB of RAM. Data is captured via a depth sensor connected via a high-speed USB-C interface to ensure stable data transmission. The software environment runs on a 64-bit operating system architecture optimized to maximize hardware processing capabilities.

The implementation of the RTAB-Map commonly uses D435i and as written in the paper [13]. The research proves the performance of D435i in indoor environment. However, in this research used D455i model as a replacement for D435i. the transition to D455i have taken some advantage of the depth estimation compared to the D435i, the graph result from the paper is shown in

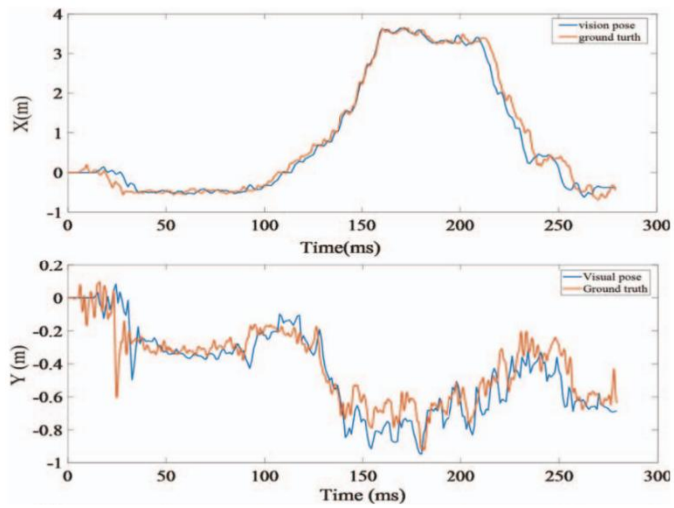

Fig 3. result from D435i

B. Data Flow

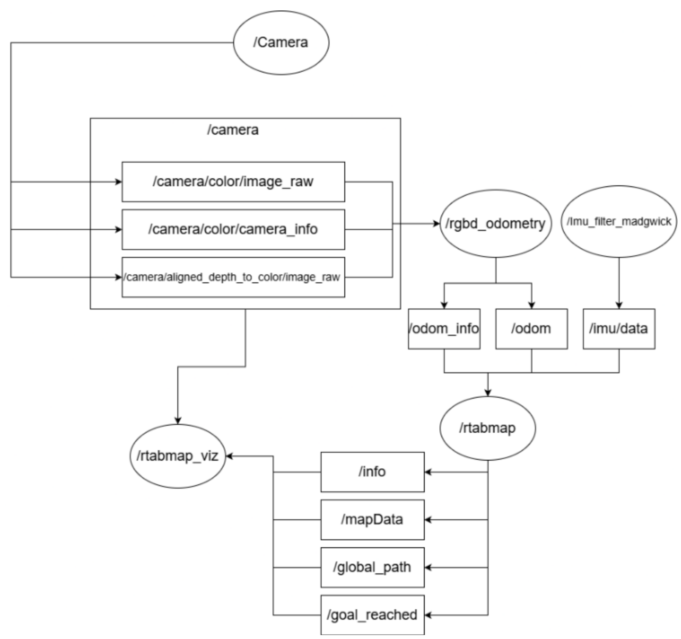

Fig 4. Block diagram for data flow of RTAB-Map

In the Fig 4 represent data flow in the RTAB-Map system with Intel RealSense D455i and IMU inside ROS2 environment. Start from the */camera* node that already available in ROS2 plugin to run RealSense camera as the main source of visual perception, which publish raw image through topic */camera/color/image_raw* like in Fig 1, and the camera intrinsic are obtain from topic */camera/color/camera_info* , as well as depth images that been aligned with color images via

/camera/aligned_depth_to_color/image_raw. These three topics are then utilized by the */rgbd_odometry* node to perform RGB-D based odometry estimation as shown in Fig 5, which calculates continuous changes in robot pose based on visual and depth information.

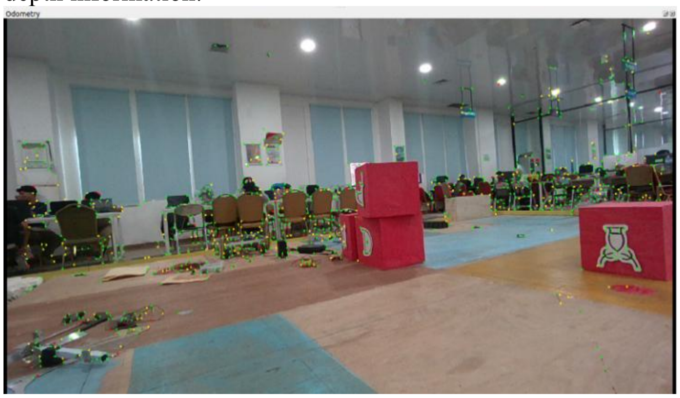

Fig 5. Visual Odometry

Inside the Fig 5 there is some dot that indicate the anchor point, this anchor point that can detect whether the position is change or not with green dots indicate inliers, yellow mean not matched features from previous frame (s), and red dot indicate outliers. Also, there is a background color for this one dark red indicate odometry lost and dark yellow indicate low inliers. To be precisely inliers are data points (like feature matches, pixels, or observations) that conform to the underlying pattern or model (e.g., correct feature matches for object recognition, normal scene pixels, and outliers conversely, are abnormal or incorrect data points that deviate significantly from the expected distribution, such as mismatched features in image stitching or noisy sensor readings, which can severely degrade the performance of vision algorithms.

In other window there is also loop closure to verify whether the robot has been in this position or not as shown in Fig 6.

Fig 6. Loop closure detection and match id

The purpose is to avoid drifting from the starting point, so when the visual match the image with the start it automatically aligns the robot position back to the correct one, and so the drift will be so minimal

In parallel running, inertial data from the IMU that actually obtain from the camera and processed by the */imu_filter_madgwick* node to produce a more stable orientation estimate in the /imu/data topic. Odometry topic which is */odom* is topic that contain the odometry data of that have been process by */rgbd_odometry* then passed to the core node */rtabmap*, which is responsible for simultaneous mapping and localization. This node combines visual, depth, and inertial data to build a map of the environment and estimate the robot's global position. The results of RTAB-Map processing are then published in various topics, such as */mapData* and */info*, which can be visualized through the */rtabmap_viz* node, as well as topics related to planning and navigation such as */global_path* and */goal_reached* status.

III. RESULT AND DISCUSSION

In this paper, the robot position obtained by D455i is transmitted to the RTAB-Map system through ROS, so as to display the performance of the D455i and RTAB-Map positioning system.

D455i is place right in front of the robot in order to easily see the environment. D455i uses the Visual Inertial Odometry with built in IMU sensor inside the D455i and through RTAB-Map, which can obtain the position information and angle direction of the robot. The position result is shown in the Fig 7.

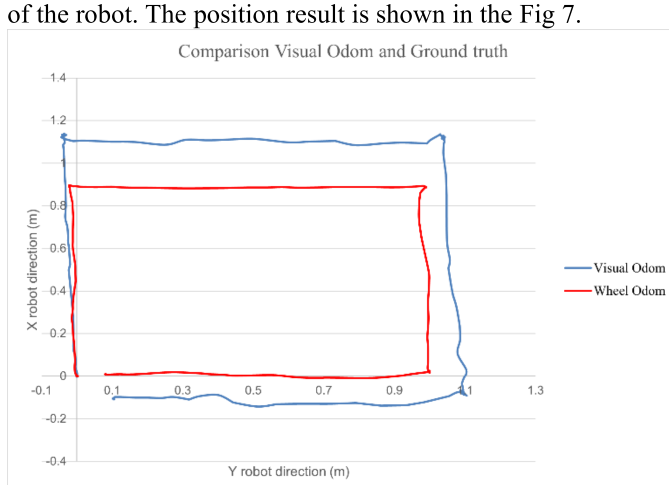

Fig 7. Error Comparison Visual Odom (blue line) and the ground truth (red line)

TABLE I
ERROR BETWEEN VISUAL ODOMETRY AND GROUND TRUTH

	X	Y
Maximum Error	0.28 m	0.18 m
Average Error	0.051 m	0.024 m
Minimum Error	0 m	0 m

A comparative analysis between the estimated visual odometry trajectory and ground truth is presented visually in Fig 7. During the initialization phase, both trajectories show a very high level of accuracy, indicating that the system is capable of estimating the initial pose with a localization error close to zero. However, as the distance traveled increases, significant spatial deviations begin to be observed, with the accumulated position error peaking at 0.28 meters.

This temporary degradation in precision can be attributed to several technical factors, including uncertainty in perception due to variations in light intensity, a decrease in the density of visual features in certain environments, or noise interference in the inertial sensor. Despite this drift or deviation, the system demonstrates robust recovery capabilities. As the mapping process progresses, the estimated trajectory gradually converges with the ground truth trajectory. This self-correction capability proves the effectiveness of the global optimization or loop closure mechanism in the Visual SLAM framework used, which is able to dynamically reduce error accumulation. This behavior pattern confirms that the system has sufficient resilience to maintain long-term localization accuracy in challenging environmental conditions.

The accuracy and reliability of the estimated pose are further demonstrated through the generation of high-precision three-dimensional reconstructions. This reconstruction, processed through the RTAB-Map (Real-Time Appearance-Based Mapping) framework, is shown in Fig 8. The visual integrity of the resulting point cloud or mesh serves as a qualitative benchmark, indicating that the system successfully maintains global consistency. The absence of significant ghosting effects or misalignment in the 3D model confirms that the visual odometry and loop closure mechanisms function effectively during the mapping process.

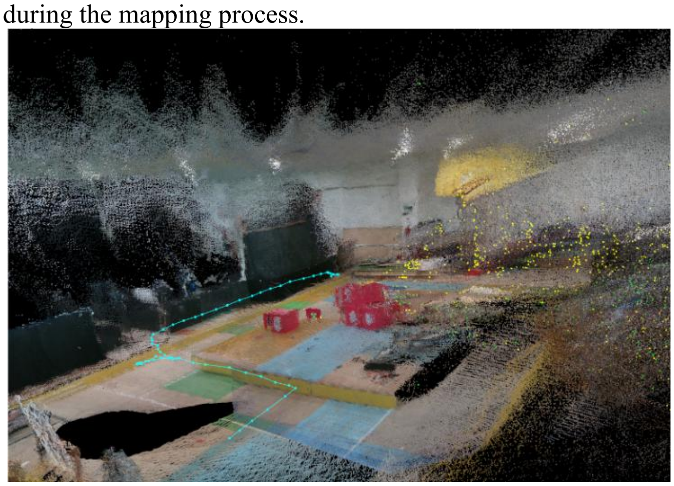
Fig 8. 3d reconstruct map

The generated map offers a spatial depiction of the surroundings, integrated with the robot's movement path. Fig 8 illustrates the robot's historical path with a cyan line, which demonstrates good congruence with the developed 3D structure. This congruence suggests that the pose estimation was both consistent and precise during the mapping operations. The uniform smoothness of the 3D reconstruction further implies consistent localization, free from notable deviations or abrupt changes. A minor area not accounted for, appearing as a shadowed region close to where the robot began, is discernible on the map. This anomaly arises from the robot's failure to capture data from that particular area at the commencement of its operation. Nevertheless, this constraint is permissible for its intended use, given that the system is engineered for Stage 2 deployment. During this stage, the robot's wheels are elevated, rendering comprehensive ground coverage unnecessary.

For another comparison there is also difference between the result from D435 as a reference point

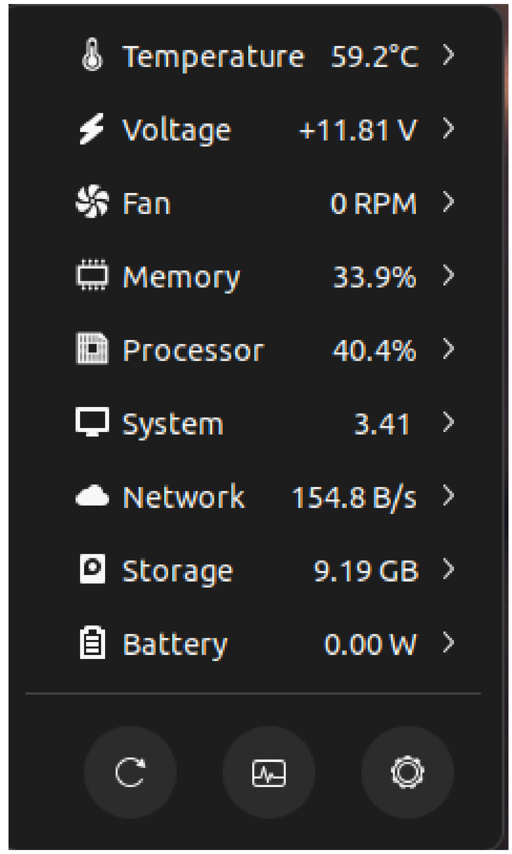

Fig 9. Cost while running

In terms of computing performance, the system shows a CPU usage rate of 40%. This figure is considered optimal because it provides sufficient resources for the next development phase. In the second stage, the robot is planned to run an object detection algorithm to identify box-shaped targets in real time. Given that the object detection algorithm has a high computational load, the current CPU usage efficiency is very important to ensure system stability and prevent processing congestion. For more detail is shown in Fig 8, the RAM consumption is around 33.9%, and the CPU consumption is around 40.4%.

IV. CONCLUSION

In this paper, the robot is equipped with an Intel RealSense D455i sensor and utilizes RTAB-Map for both odometry estimation and simultaneous localization and mapping within the ROS environment. The integration of the RealSense camera with the RTAB-Map framework, supported by the ROS-provided RealSense launch configuration and the imu_filter_madgwick node, enables effective fusion of visual, depth, and inertial data. This sensor fusion significantly improves the stability, accuracy, and efficiency of robot pose estimation, while also enhancing the quality of the generated 3D reconstruction. The experimental results demonstrate that the proposed system is capable of producing a consistent and smooth 3D map, with the estimated robot trajectory closely aligned with the reconstructed environment. The continuity of the robot path and the absence of severe distortions or discontinuities in the 3D reconstruction indicate that localization drift is well controlled. These results confirm that the RTAB-Map–based RGB-D SLAM approach, combined with IMU filtering, provides reliable performance for indoor mapping scenarios. Overall, the obtained results validate that the proposed system achieves accurate robot positioning and high-quality 3D reconstruction, making it suitable for practical robotic applications that require robust localization and mapping.

REFERENCES

[1] M. Kulkarni, P. Junare, M. Deshmukh, and P. P. Rege, "Visual SLAM Combined with Object Detection for Autonomous Indoor Navigation Using Kinect V2 and ROS," in *2021 IEEE 6th International Conference on Computing, Communication and Automation, ICCCA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 478–482. doi: 10.1109/ICCCA52192.2021.9666426.

[2] G. Huang, L. Lu, Y. Zhang, G. Cao, and Z. Zhou, "Improved vision-only localization method for mobile robots in indoor environments," *Autonomous Intelligent Systems*, vol. 4, no. 1, Dec. 2024, doi: 10.1007/s43684-024-00075-9.

[3] *EST 2019 Colchester : 2019 Eighth International Conference on Emerging Security Technologies (EST) : 22-24 July, 2019, Colchester, Essex, UK*. IEEE, 2019.

[4] D. U. Rijalusalam and I. Iswanto, "Implementation kinematics modeling and odometry of four omni wheel mobile robot on the trajectory planning and motion control based microcontroller," *Journal of Robotics and Control (JRC)*, vol. 2, no. 5, pp. 448–455, Sep. 2021, doi: 10.18196/jrc.25121.

[5] A. K. Gharehbagh, R. Judeh, J. Ng, C. Von Reventlow, and F. Röhrbein, "Real-time 3D Semantic Mapping based on Keyframes and Octomap for Autonomous Cobot," in *2021 9th International Conference on Control, Mechatronics and Automation, ICCMA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 33–38. doi: 10.1109/ICCMA54375.2021.9646203.

[6] C. J. Lin, C. C. Peng, and S. Y. Lu, "Real-Time Localization for an AMR Based on RTAB-MAP," *Actuators*, vol. 14, no. 3, Mar. 2025, doi: 10.3390/act14030117.

[7] X. Zou *et al.*, "Novel Visual Odometry Method for Water Autonomous Navigation," *2022 IEEE 2nd International Conference on Computation, Communication and Engineering, ICCCE 2022*, pp. 13–17, 2022, doi: 10.1109/ICCCE55785.2022.10036173.

[8] B. N. Y. Divan, M. J. Iorpenda, and V. Willert, "Getting Started with a Simple Visual-Inertial Odometry," *2024 22nd International Conference on Research and Education in Mechatronics, REM 2024*, pp. 231–236, 2024, doi: 10.1109/REM63063.2024.10735689.

[9] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," Aug. 2017, doi: 10.1109/TRO.2018.2853729.

[10] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *J Field Robot*, vol. 36, no. 2, pp. 416–446, Mar. 2019, doi: 10.1002/rob.21831.

[11] P. Hübner, J. Hou, and D. Iwaszczuk, "EVALUATION OF INTEL REALSENSE D455 CAMERA DEPTH ESTIMATION FOR INDOOR SLAM APPLICATIONS," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Dec. 2023, pp. 1207–1214. doi: 10.5194/isprs-archives-XLVIII-1-W2-2023-1207-2023.

[12] X. Wang, L. Ma, D. Wang, and B. Su, "Autonomous UAV-Based Position and 3D Reconstruction of Structured Indoor Scene," in *Proceeding - 2021 China Automation Congress, CAC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 3914–3919. doi: 10.1109/CAC53003.2021.9727567.

[13] G. Ramirez, A. Montes De Oca, and G. Flores, "3D maps of vegetation indices generated onboard a precision agriculture UAV," in *2023 International Conference on Unmanned Aircraft Systems, ICUAS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 564–571. doi: 10.1109/ICUAS57906.2023.10156582.

[14] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021, doi: 10.1109/TRO.2021.3075644.