Mapping and Localization in 3D Space for Vision-based Robot Manipulation

Md Tanzil Shahria and Md Samiul Haque Sunny
Department of Computer Science
University of Wisconsin-Milwaukee
Wisconsin, United States
mshahria@uwm.edu, msunny@uwm.edu

Md Ishrak Islam Zarif and Sheikh Iqbal Ahamed
Department of Computer Science
Marquette University
Wisconsin, United States
ishrak.zarif@marquette.edu, sheikh.ahamed@marquette.edu

Mohammad H. Rahman
Department of Mechanical Engineering
University of Wisconsin-Milwaukee
Wisconsin, United States
rahmanmh@uwm.edu

Abstract

As many factors that can be attributed to a high-performance robotic system and with increasing demand for robotic application in automation or as assistive devices, research in this area has received much attention. Robot manipulation is one of the vital components of the industrial or assistive robotic system. The utilization of various vision-based approaches and different associated algorithms enhances the overall robot manipulation system's performance, functionality, and efficiency. Implementing a robust and accurate vision-based robot manipulation system is still today's challenge in robotics applications because of sensing and environmental uncertainties. In this paper, a computer vision-based object detection and localization system for robot manipulation system is proposed, which enables the robot to locate any object in the 3D space. The system uses a set of RGB camera, depth sensor, and vision processor that allows the system to get the coordinates of any objects in 3D space in terms of the camera to map and localize the object. Additionally, data associates are compared by statistical distance. Finally, a series of systematic experiments are performed to verify the reliability and accuracy of the proposed system. The system can successfully detect different objects and calculate data from the image frame with an accuracy of 94%.

Keywords
Robot Manipulation, Computer Vision, Mapping, Localization, Depth Sensor

1. Introduction
A robot that has intelligent control should be able to perceive its environment and interact with it (He, Li, and Chen 2017). Among the essential abilities, the ability to manipulate objects is fundamental and significant in that it will bring enormous power to society. For example, industrial robots can accomplish the pick-and-place task laborious for human laborers, and domestic robots can assist disabled or older people in their daily grasping tasks (Chiacchio, Petropoulos, and Pichler 2018; Martinez-Martin and Del Pobil 2017). Increasing the perception capabilities of robots has been a long-standing goal of computer vision and robotics. Object detection and pose estimation are some of the

essential functions of an autonomous robot, especially when the robot needs the ability to search, measure, and manipulate objects while interacting with humans more intelligently.

A robot needs to have a basic set of capabilities to perform physical tasks (Ben-Ari and Mondada 2018). It must be physically capable of doing its job, which includes mobility and manipulation. It must be able to perceive its environment sufficiently well to accomplish its task. In most cases, the robot must build a map of its environment by making measurements with its sensors and estimating its current location (Mišeikis et al. 2016). Finally, it must be able to reason about its mobility and manipulation and its perception of the local environment to perform its task efficiently and safely (Hebert et al. 2015).

The mapping, localization, and planning are problems that lie at the core of robot manipulation (Lösch et al. 2018; Ollero and Siciliano 2019; Frank, Moorhead, and Kapila 2016). Simultaneous Localization and Mapping (SLAM) has interdependency between mapping and localization. To a large degree, researchers have been solving for small, two-dimensional, structured environments. To make robots generally useful in the broader world, they need to move beyond simple environments into large, three-dimensional, and potentially unstructured ones. Accordingly, they need general algorithms for mapping, localizing, planning, and exploring that work just about anywhere. This study is concerned with developing broadly applicable methods that will allow robots to explore and operate in most environments, including indoor and outdoor, flat, and highly three-dimensional.

We make several assumptions about what is involved in a general approach:

1. We believe that the approach must be constant in computation (real-time) and linear in storage space to fit within the computational constraints of real mobile robots.
2. We believe that map representations must be fully 3D and capable of representing arbitrary 3D geometry at a level of resolution that is appropriate for the robot and its task.
3. While the map needs to be locally accurate for path planning and obstacle avoidance, global accuracy can often be relaxed, again depending on the robot and its task.
4. We believe that the method must not rely on any structures or features in the environment.

As a result, we believe that the approach must use metric range measurements and succeed even when these measurements are sparse and inaccurate.

In this paper, the goal is to create a system for assistive robots for detection, pose estimation, and measurement of an object. Thus, it will be useful to create conditions for the execution of tasks such as manipulation of the object

The rest of this paper is organized as follows: Section 2 discusses some literature review in this field, Section 3 presents the method used in this study, Section 4 illustrates the experimental setup of the overall experiment, Section 5 represents the result of this study, and finally, Section 6 draws the conclusion of the study.

2. Literature Review

In vision-based robot manipulation, to find the objects from a scene, classify them, and measure their location, several approaches were proposed. Roughly we can divide them into two categories: model-based, and appearance-based (Lins, Givigi, and Kurka 2015). Features like edges, straight lines, contours, and segmentation are used in model-based approaches to compare the image with a 2D model or projections of a 3D model of the objects. Global shape, statistical characteristics, and local features are used in appearance-based approaches. A physical task in robotic manipulation requires scene interpretation in the unconstructed environment for two types of sensors: contact sensors and contactless sensors (Mateo, Gil, and Torres 2016). Force and tactile sensors are counted as contact sensors, while contactless sensors include visual sensors based on images. Many studies combined data from both types of sensors to perform tasks in the unknown environment, including motion planning, grasping, and intelligent manipulation of objects.

A simultaneous kinematic calibration, localization, and mapping are proposed which can calibrate the kinematic parameters of an industrial robot manipulator using a commercial RGB-D camera attached to its end effector to reconstruct the surroundings (Li et al. 2019). Feature detection and geometry are used to get the kinematic calibration. The application and research progress of harvesting robots and vision technology in localization, target recognition, 3D reconstruction, and fault tolerance of complex agricultural environment is presented in (Tang et al. 2020). Estimation of the localization of the objects in 3D scenes using collaborative robots is done by (Lins, Givigi, and Kurka 2015). Vision-based object manipulation for construction application using two stereo cameras and a robotic arm mounted on a mobile platform (Asadi et al. 2021). Pose estimation using RGB-based and RGB-D-based methods is mentioned in (Du, Wang, and Lian 2019) for vision-based robotic grasping. In (Zhao et al. 2021) two-stage

positioning and object pose estimation in the robotic system are presented, which includes the vision-based rough positioning stage and tactile-based precise positioning stage.
The real environments are generally chaotic and unstructured, the posture of the object to be processed in the environment appears randomly, and the robotic arm cannot process objects with random poses (Cheng, Yen, and Jeng 2021). Therefore, researchers agreed that robotic applications either in construction, industry, or in assistive tasks for individuals with disabilities need computer vision and other sensing mechanisms to supply the robot with real-time data and information of its physical environment (Giftthaler et al. 2017; Wu, Jiang, and Song 2015; Feng et al. 2015). In many of the projects described above, the computer vision system perceives the information of the working environment through the image sensor, thereby processing and analyzing the environmental information, allowing the robots to complete more complex processing operations autonomously.

3. Methods

To detect and get coordinates of an object in the 3D environment, the proposed system uses the ArUco detector ('OpenCV: Detection of ArUco Markers' 2021) and OpenCV. Using this approach, the system can detect any object within the image frame and generate the x and y coordinates of the center of that object along with its width and height. The coordinates will help the overall system to localize that particular object. On the other hand, the width and height of the object will give additional information to the system for different tasks and further analysis. To localize the object in the 3D space, a third coordinate (depth) is needed, which can be generated using the depth channel of the RealSense camera ('Depth Camera D435' 2021). Using the 2D coordinate of the center of any object and the depth frame, the system calculates the distance of that object from the camera and uses it as the z-coordinate of the object.

3.1 System Architecture

At first, the model loads the ArUco detector and initializes the object detector model and the RealSense camera. If the camera is connected, the model takes the RGB frame and depth frame from the RealSense camera as input.
To make this model work, there should be an ArUco marker in the input frame. Figure 1 presents the ArUco marker used in this study. The ArUco detector uses this marker as a reference, and by reading the marker, the ArUco detector initializes the parameter. The model uses this parameter to detect corners (borderlines) of different objects from the RGB frames.

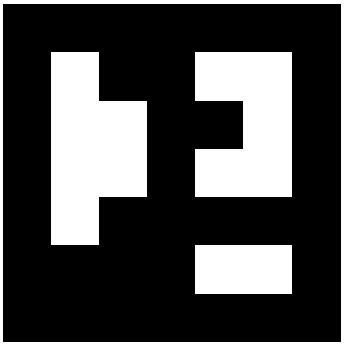

Figure 1. ArUco marker used in this study

Using OpenCV, the model creates different rectangles around objects and calculates the height and width of each object using the marker parameters. After that, the system identifies the x and y coordinates of the center point of different objects in terms of pixel values. Then the model uses the depth frame and the x, y coordinates of each object to calculate the distance of that particular object from the camera. After all the calculations, the model displays the RGB frame and sends the x, y, and distance value to the next step as a 3D coordinate. It also passes the height and width of different objects to the next step along with the coordinates. Figure 2 presents the workflow of the proposed system.

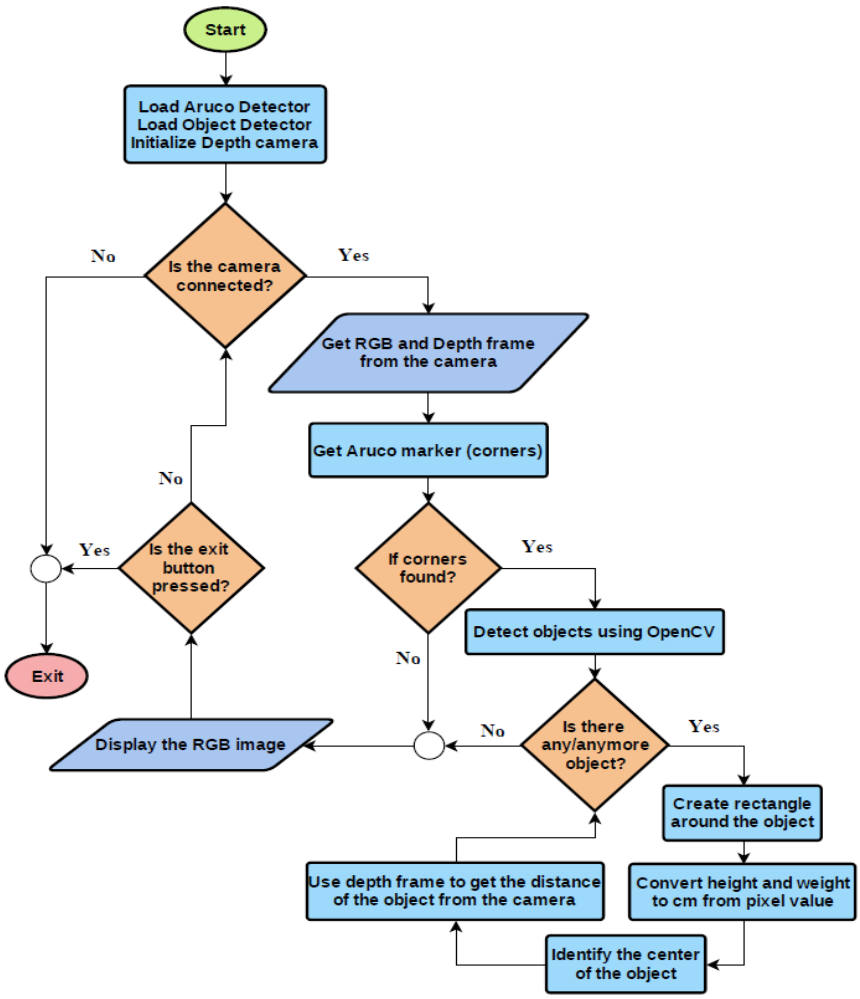

Figure 2. Flowchart of the proposed OpenCV-based algorithm

There are only three libraries used in this system: OpenCV, pyrealsense2, numpy. The pseudo-code of the proposed system is presented below:

```
SET parameters TO OpenCV_Aruco_Detector
SET detector TO ObjectDetectorModel

WHILE True:
   GET RGB_frame, depth_frame FROM Camera

   IF corners:
      CREATE border_lines around the object
      SET contours TO detect_objects USING detector and OpenCV_Aruco_Detector FROM image

      FOR cnt IN contours:
         SET (x, y), (w, h), angle TO rectangle_box
         SET object_width TO pixel_cm_ratio
         SET object_height TO pixel_cm_ratio
         SET OUTPUT (x, y, z, width, height) as Text on Detected image

   DISPLAY image_frame until EXIT
```

4. Experimental Setup

For completion of this work, we are using the NVIDIA Jetson Nano Developer Kit 4GB version for computation and detection purposes and Intel RealSense Depth Camera D435 for capturing the RGB as well as depth frame from the image.

The NVIDIA Jetson Nano is a development kit and embedded system-on-module (SoM) from the NVIDIA Jetson lineup ('Jetson Nano Developer Kit' 2021). It is built in a small form factor and has powerful computational power, making it ideal for computer vision and deep learning applications. Jetson Nano has a 128-core Maxwell GPU, quad-core ARM A57 64-bit CPU, 4GB LPDDR4 memory, and MIPI CSI-2 and PCIe Gen2 high-speed I/O capabilities. Jetson Nano uses the NVIDIA JetPack SDK to run Linux and gives 472 GFLOPS of FP16 computation performance while consuming 5 to 20W of electricity.

The Intel RealSense Depth Camera D435 features four lenses, an RGB module for regular photo and video, along with an IR projector and a right imager, as well as a left imager for depth sensing ('Depth Camera D435' 2021). This USB-powered camera calculates depth data using the stereo depth method. Intel RealSense Depth Camera D435 offers accurate depth perception when the object is moving, or the device is in motion with the global image shutter and wide field of view. Two depth sensors are spaced a short distance apart in the camera module. A stereo camera compares the two images produced by these two sensors. These comparisons provide depth information because the distance between the sensors is known.

The RealSense camera captures videos of the object placed in front of it and sends the video data to the Jetson Nano. Then the Jetson Nano split frames from the video and started processing the image. After the computation and detection of objects, it calculates coordinates, distance value, weight, as well as height information, and places on the frame. Then it outputs the results to the connected display. Figure 3 illustrates the experimental setup of mapping and localization in 3D space.

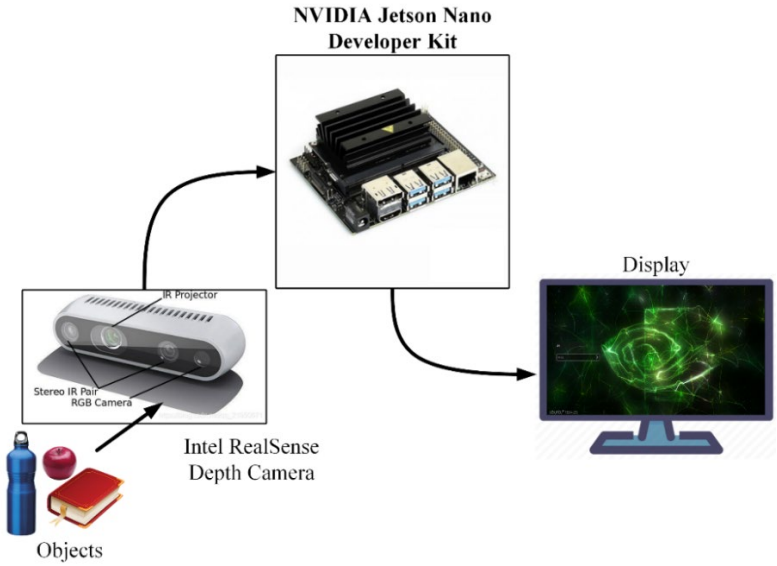

Figure 3. Experimental setup of mapping and localization in 3D space

5. Results and Discussion

In the experiment, we used NVIDIA Jetson Nano as the development kit and Intel RealSense Depth Camera as the input device. The proposed model successfully detects different objects from the image frame. It can accurately measure the 3D coordinates of the center of different objects with respect to the center point of the camera along with the height and width of that object. The overall accuracy of the proposed system is 94%. The accuracy is calculated by validating the coordinates, height, and width of different objects. Figure 4 presents the detection of 3D coordinates along with the height and width of different objects in real-time.

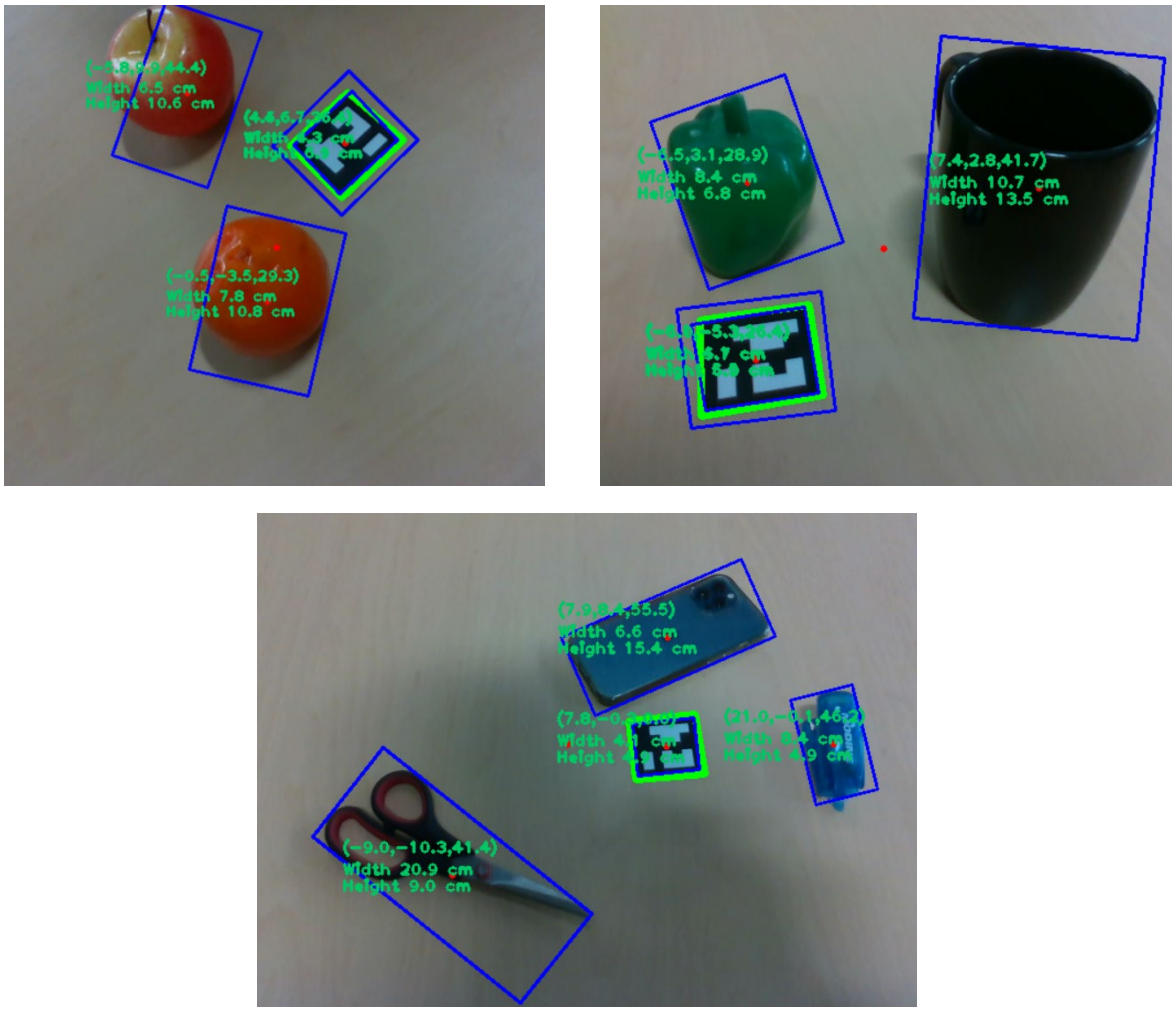

Figure 4. 3D coordinates along with height and width of different objects

Although the proposed model works really well with the environment, there are still some limitations to the model. To detect and calculate different data of the objects, the ArUco marker has to be in the frame, which is not feasible in some cases. In the future, we will explore other approaches with object recognition and design a complete robust model for different applications.

6. Conclusion

Robot manipulation is one of the fundamental segments in the field of robotics. To assist the robot in this task, various vision-based approaches are being explored by the researchers. In this research, we have enlightened a computer vision-based approach to detect and localize different objects in 3D space for a robotic system. We have used the ArUco detector and OpenCV to detect any object within the image frame and generate the x, y, and z coordinates of the center of that object along with its width and height. These data help the overall system to localize any particular object in the 3D environment. The RGB channel and depth channel of the RealSense camera are used as the input source, and Jetson Nano is used as the development kit in this study. Lastly, a series of methodical experiments are performed to verify the overall performance of the proposed system. The system can successfully detect different objects, generate 3D coordinates, and calculate the height and width of each object with an accuracy of 94%. In the future, we will continue exploring other possible approaches to design a more robust model for different robotic applications.

Acknowledgment:

This material is based upon work supported by NASA under Award No. RIP23_1-0 issued through Wisconsin Space Grant Consortium. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

References

Asadi, Khashayar, Varun R Haritsa, Kevin Han, and John-Paul Ore. 2021. 'Automated Object Manipulation Using Vision-Based Mobile Robotic System for Construction Applications', Journal of Computing in Civil Engineering, 35: 04020058.

Ben-Ari, Mordechai, and Francesco Mondada. 2018. 'Robots and Their Applications.' in Mordechai Ben-Ari and Francesco Mondada (eds.), Elements of Robotics (Springer International Publishing: Cham).

Cheng, Fang-Che, Chia-Ching Yen, and Tay-Sheng Jeng. 2021. 'Object Recognition and User Interface Design for Vision-based Autonomous Robotic Grasping Point Determination'.

Chiacchio, Francesco, Georgios Petropoulos, and David Pichler. 2018. "The impact of industrial robots on EU employment and wages: A local labour market approach." In.: Bruegel working paper.

'Depth Camera D435'. 2021. https://www.intelrealsense.com/depth-camera-d435.

Du, Guoguang, Kai Wang, and Shiguo Lian. 2019. 'Vision-based robotic grasping from object localization, pose estimation, grasp detection to motion planning: A review', arXiv preprint arXiv:1905.06658, 2.

Feng, Chen, Yong Xiao, Aaron Willette, Wes McGee, and Vineet R. Kamat. 2015. 'Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites', Automation in Construction, 59: 128-38.

Frank, Jared A, Matthew Moorhead, and Vikram Kapila. 2016. "Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks." In 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 302-07. IEEE.

Giftthaler, Markus, Timothy Sandy, Kathrin Dörfler, Ian Brooks, Mark Buckingham, Gonzalo Rey, Matthias Kohler, Fabio Gramazio, and Jonas Buchli. 2017. 'Mobile robotic fabrication at 1: 1 scale: the in situ fabricator', Construction Robotics, 1: 3-14.

He, Wei, Zhijun Li, and CL Philip Chen. 2017. 'A survey of human-centered intelligent robots: issues and challenges', IEEE/CAA Journal of Automatica Sinica, 4: 602-09.

Hebert, Paul, Max Bajracharya, Jeremy Ma, Nicolas Hudson, Alper Aydemir, Jason Reid, Charles Bergh, James Borders, Matthew Frost, and Michael Hagman. 2015. 'Mobile manipulation and mobility as manipulation—design and algorithms of RoboSimian', Journal of Field Robotics, 32: 255-74.

'Jetson Nano Developer Kit'. 2021. https://developer.nvidia.com/embedded/jetson-nano-developer-kit.

Li, Jinghui, Akitoshi Ito, Hiroyuki Yaguchi, and Yusuke Maeda. 2019. 'Simultaneous kinematic calibration, localization, and mapping (SKCLAM) for industrial robot manipulators', Advanced Robotics, 33: 1225-34.

Lins, R. G., S. N. Givigi, and P. R. G. Kurka. 2015. 'Vision-Based Measurement for Localization of Objects in 3-D for Robotic Applications', IEEE Transactions on Instrumentation and Measurement, 64: 2950-58.

Lösch, Robert, Steve Grehl, Marc Donner, Claudia Buhl, and Bernhard Jung. 2018. "Design of an autonomous robot for mapping, navigation, and manipulation in underground mines." In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1407-12. IEEE.

Martinez-Martin, Ester, and Angel P Del Pobil. 2017. 'Object detection and recognition for assistive robots: Experimentation and implementation', IEEE Robotics & Automation Magazine, 24: 123-38.

Mateo, CM, P Gil, and F Torres. 2016. 'Visual perception for the 3D recognition of geometric pieces in robotic manipulation', The International Journal of Advanced Manufacturing Technology, 83: 1999-2013.

Mišeikis, Justinas, Kyrre Glette, Ole Jakob Elle, and Jim Torresen. 2016. "Multi 3D camera mapping for predictive and reflexive robot manipulator trajectory estimation." In 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 1-8. IEEE.

Ollero, Anibal, and Bruno Siciliano. 2019. Aerial robotic manipulation (Springer).

'OpenCV: Detection of ArUco Markers'. 2021. https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html.

Tang, Yunchao, Mingyou Chen, Chenglin Wang, Lufeng Luo, Jinhui Li, Guoping Lian, and Xiangjun Zou. 2020. 'Recognition and Localization Methods for Vision-Based Fruit Picking Robots: A Review', Frontiers in Plant Science, 11.

Wu, C., S. Jiang, and K. Song. 2015. "CAD-based pose estimation for random bin-picking of multiple objects using a RGB-D camera." In 2015 15th International Conference on Control, Automation and Systems (ICCAS), 1645-49.

Zhao, Dan, Fuchun Sun, Zongtao Wang, and Quan Zhou. 2021. 'A novel accurate positioning method for object pose estimation in robotic manipulation based on vision and tactile sensors', The International Journal of Advanced Manufacturing Technology, 116: 2999-3010.

Biographies

Md Tanzil Shahria received the B.S. degree in Computer Science and Engineering from the North South University, Bangladesh in 2018. He is a Ph.D. student at the BioRobotics Lab, University of Wisconsin-Milwaukee, USA. He has been researching in different learning-based systems since 2019. His work focuses on computer vision, vision-based manipulation, real-time object localization, and robot navigation. He is currently working on developing a vision-based manipulation system for assistive robot using learning-based approach.

Md Ishrak Islam Zarif is currently working as a research assistant under the ubicomp lab of Marquette University as well as the Biorobotics Lab of the University of Wisconsin-Milwaukee. His research interest on telerehabilitation system development. At present, he is pursuing his Ph.D. from the Department of Computer Science at Marquette University, USA. He has a background working in data science, machine learning, IoT, and system development. He completed his undergrad at Khulna University of Engineering & Technology, Bangladesh, in 2018. Mr. Zarif has publications in his research area in different renowned conferences and journals.

Md Samiul Haque Sunny was born in Netrokona, Bangladesh in 1994. He received the B.S. degree in Electrical and Electronic Engineering from the Khulna University of Engineering and Technology in 2017. He is a Ph.D. student at the BioRobotics Lab, University of Wisconsin-Milwaukee, with a background in Artificial Intelligence, digital signal and image processing, data mining, robotics, biological signal processing, human-machine interface design, and power system stability. He is currently working on developing an eye-gaze controlled user interface deployable to Hololnes-2's mixed-reality platform to enable a collaborative work environment for individuals with limited upper limb movement, EEG signal for better BCI application, and structures of CNN for upgrading its performance in image recognition.

Sheikh Iqbal Ahamed is Professor and Chair at the Department of Computer Science and Director of the Ubicomp Research Lab. He received his PhD in Computer Science from the Arizona State University in 2003 under the direction of Dr. Stephen S. Yau. Dr. Ahamed has a strong performance record in research, with a career record of over $1 million in external research funding. His current projects cover a number of state-of-the-art advances in medical computing and mobile computing. Dr. Ahamed's research work focuses on building customized and innovative patient communication methods through technology, developing new and innovative approaches for health monitoring, pain management, mapping technologies, and activity monitoring for smartphones. He has worked with hospitals in the U.S. and internationally on a number of projects, as well as with leading healthcare companies in the healthcare industry. Dr. Ahamed has worked with a number of engineers, nurses, and physicians on 20 healthcare grants over the past 13 years. Projects included work with cellphones, sensors, tablets, web applications, and HIPAA compliant cloud servers. His experience extends to working with patient populations in American Indian and Hispanic communities.

Mohammad H Rahman is with the Mechanical and Biomedical Engineering Department, University of Wisconsin-Milwaukee, WI, USA. As Director of the BioRobotics Lab at the University of Wisconsin-Milwaukee, he brings the resources and expertise of an interdisciplinary R&D team. For more than 15 years he has been researching bio-mechatronics/bio-robotics with emphasis on the design, development and control of wearable robots to rehabilitate and assist elderly and physically disabled individuals who have lost their upper-limb function or motion due to stroke, cardiovascular disease, trauma, sports injuries, occupational injuries, and spinal cord injuries. He received a BSc Engineering (mechanical) degree from Khulna University of Engineering & Technology, Bangladesh in 2001, a Master of Engineering (bio-robotics) degree from Saga University, Japan in 2005 and a PhD in Engineering (bio-robotics) from École de technologie supérieure (ETS), Université du Québec, Canada in 2012. He worked as a postdoctoral research fellow in the School of Physical & Occupational Therapy, McGill University (2012-2014). His research interests are in bio-robotics, exoskeleton robot, intelligent system and control, mobile robotics, nonlinear control, control using biological signal such as electromyogram signals.