Design of a voice control 6DoF grasping robotic arm based on ultrasonic sensor, computer vision and Alexa voice assistance

Zhiheng Wang

Department of Electrical and Electronics Engineering

London South Bank University

London, UK

[wzhlove2003@gmail.com](mailto:wzhlove2003@gmail.com)

Daqing Chen

Department of Electrical and Electronics Engineering

London South Bank University

London, UK

chend@lsbu.ac.uk

Perry Xiao

Department of Electrical and Electronics Engineering

London South Bank University

London, UK

xiaop@lsbu.ac.uk

*Abstract*— The article presents a study to design a 6-degree of freedom robotic arm which can pick up objects in random positions on a 2D surface based on Arduino microcontroller, ultrasonic sensors and picamera. The robotic arm is able to recognise objects based on computer vision algorithm for shape detection. The ultrasonic sensor measures the distance between the objects and the robotic arm, and the position of the objects in the real world will be detected by its mass centre in the image to improve the accuracy of the pick-up movement. Arduino microcontroller will calculate the rotation angles for the joints of the robotic arm by using inverse kinematics algorithms. The movement of the robotic arm also can be controlled by an Amazon Alexa voice assistance device. The experiment of applying the artificial neural network to control the robotic arm pick-up movement is achieved. The artificial neural network can manipulate the position of the robotic arm to pick up objects after training using the values which are calculated by inverse kinematics equations. The Raspberry Pi is used for processing the computer vision data, and voice commands from Alexa Voice Service based on cloud service.

Keywords—robotic arm, computer vision, object detection, artificial neural network, Alexa voice assistance

# Introduction

Robot concepts have been introduced from science fiction[1]. As the developments of mechanical, electronics, computer, sensor and communication technologies, robots have already extensively applied in industrial manufacture and even humans' daily life. The robotic arms are one of the earliest programmable automated machines utilised in the industrial production and helped humans to finish some work faster, easier and more accurate. It is able to move objects, tools, heavy stuff and pick and place objects in automation and complete some hazardous work which human cannot do. Robotic arms have been used for a variety of applications in industry, such as welding, gripping, lifting and even automotive assembling.

Because of development of sensors, camera, microcontroller and computer, robotic arms are able to ‘feel’ outside world by sensors and ‘think’ using microcontrollers and computers, then decide what they should do. Robotic arms become very intelligent. Especially when cameras and image processing have applied into robotic arms, they can capture images by cameras and analyse images to take useful information and then process commands and take actions. Furthermore, an artificial neural network is the most modern technology in robot control. It has many advantages in robot performance especially for machine learning and deep learning. Robots are able to learn many things using the artificial neural network, such as robot movement and image recognition.

In this paper, a robotic arm is intended to pick up, move and place the objects. The location of the objects is unknown. The robotic arm can detect the desired objects and find the location of the objects by picamera and ultrasonic sensor. The inverse kinematics algorithm is applied in the robotic arm control to decide the position to grasp the objects. The experiment of applying the artificial neural network to control the pick-up movement is achieved after training.

The movement of the robotic arm also can be controlled by voice command by using an Amazon Alexa voice assistance device.

# system description

The system consists of 4 sections, the robotic arm which contains the Picamera and the ultrasonic sensor, Raspberry Pi, Arduino and motor driver shield, and Amazon Echo Dot. The schematic representation of the robotic arm control system is shown in Fig. 1.

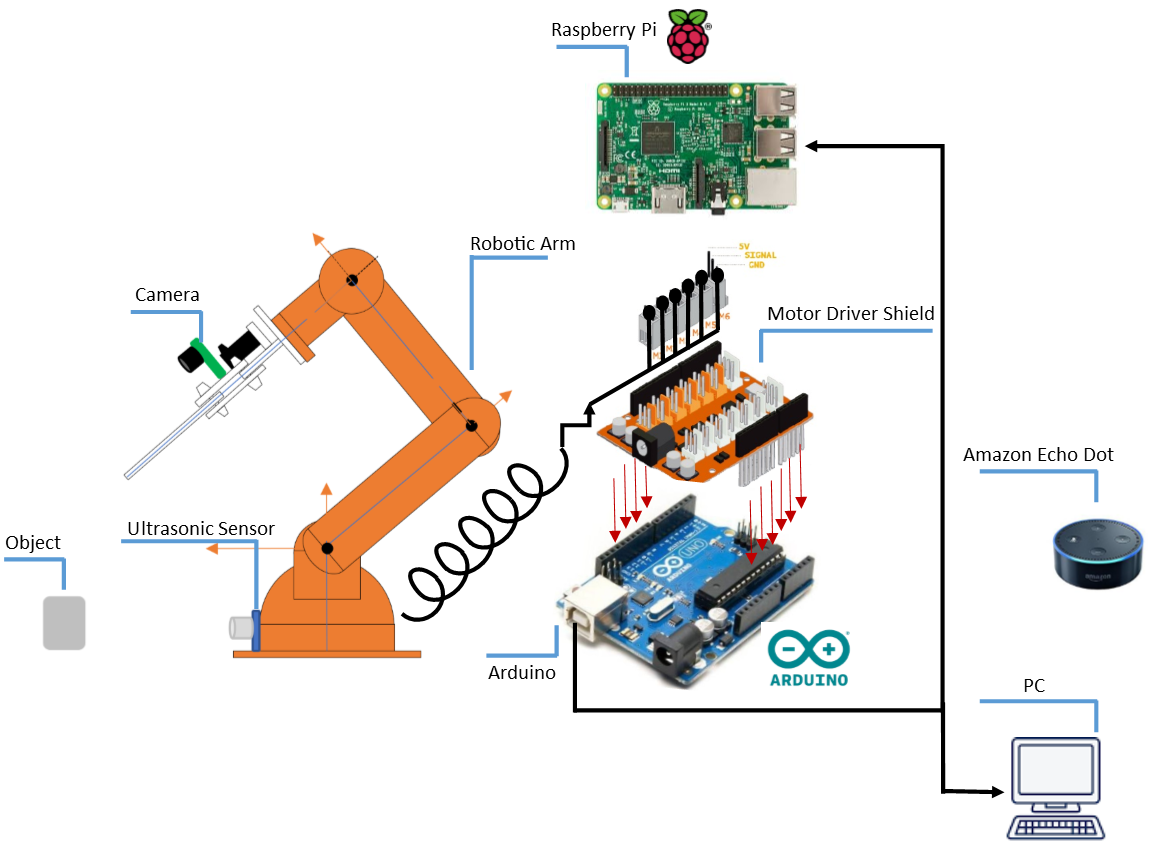


Fig. 1 The diagram of the robotic arm and control

## Robotic Arm

The robotic arm has been built by 6 servo motors, and it has 5 Degree of Freedom(DOF). It contains base, shoulder, elbow and wrist and wrist rotation joints and they are connected by its limbs. An ultrasonic sensor is installed on the base of the robotic arm in order to measure the distance between the target objects and the robotic arm. A Picamera is attached at the gripper to capture the images of the objects.

The rotation range of servo motor is 0 degree to 180 degrees. However, since the structure and balance of the robotic arm, some of the motors cannot rotate as full range, the joint’s name and rotation range are given in TABLE I.

TABLE I. The motors and rotation range of the robotic arm

|  |  |  |
| --- | --- | --- |
| Motor | Joint Name | Rotation Range |
| M1 | Base | 0° ~ 180° |
| M2 | Shoulder | 15° ~ 165° |
| M3 | Elbow | 0° ~ 180° |
| M4 | Wrist Pitch | 0° ~ 180° |
| M5 | Wrist Roll | 0° ~ 180° |
| M6 | Gripper | 10° ~ 73° (10° the gripper is open, 73° the gripper is closed. |

## Raspberry Pi and computer

The Raspberry Pi is a small single-board computer. The Raspberry Pi 3 model B is used in the system because of its fast processing speed and on-board Wi-Fi. These features are able to deal with complex computer vision algorithms and access public web service to accomplish voice control. The computer is only used to run Matlab software and simulate the artificial neural network.

## Arduino and motor driver

Arduino is a microcontroller to control the servo motors by its Pulse Width Modulation(PWM) pins and detects outside world by connecting sensors to its analogue pins. The motor driver connects the external electrical power supply and provides enough electrical current to drive servo motors.

## Amazon Echo Dot

Amazon Echo Dot is a smart speaker which is developed by Amazon. It has an array of 7 microphones and Wi-Fi chip to collect voice and access to the internet. The Echo Dot gathers the voice commands and streams audio to a cloud-based service Alexa Voice Service which is able to recognise the natural voice and interpret them, and respond the request. It also provides self-service APIs, tools, documentation, and code samples that make developers create personalised skills.

A close up of text on a white background

Description generated with high confidence

Fig. 2 The flowchart of the pickup movement

The pick-up movement which is used to pick and place the objects in a random position has applied ultrasonic distance sensor, the inverse kinematic equation of the robotic arm, Arduino control, and computer vision processing. The flowchart of the pick-up movement is shown in Fig. 2.

Once the pick-up function has been called, the ultrasonic sensor measures the distance by the echo wave. If the object is not in the range which the robotic arm can pick it up, the basement of the robotic arm then rotates for and continue to find an object in range. When the object is in the robotic arm pick-up range, then Arduino will send ‘1’ to the Raspberry Pi through serial communication. After the Raspberry Pi reads ‘1’ from the serial port, it then starts to process the video which is captured by the camera. The computer vision function is called to detect the objects and calculate the centroid of the object. If the object is not at the centre of the video frame, it means the end-effector has not pointed at the centre of the object. The Raspberry Pi will send ‘0’ to Arduino, then Arduino will control the robotic arm rotates for . When the object is at the centre of the video frame, the Raspberry Pi send ‘1’ to Arduino, and that means the end effector points at the centre of the object. Then, the robotic arm will stop rotating, and the ultrasonic sensor measures the distance between the object and the robotic arm. The distance value is used for calculating the rotation angles for shoulder, elbow and wrist motors in the robotic arm. After getting the angle values, Arduino controls the motors and moves the end-effector to the object and pick it up. After these processings, the robotic arm is able to pick up an object at any reachable location.

# kinematic model

Denavit-Hartenberg (D-H) [2] is a general transformation between attaching reference frames to the links and the links of spatial kinematic which is often used in the robotic mechanical system. The D-H method applies the four parameters which are the rotation about the z-axis, the joint offset, length of a link and twist angle[3]. The advantage of this method is the storage efficiency for dealing with the kinematics of robot chains and computational robustness[4]. The coordinate of the robotic arm and the D-H parameters are shown in Fig. 3 and TABLE II.

A picture containing text, map

Description generated with very high confidence

Fig. 3 The coordinates of the robotic arm

TABLE II D-H parameters table of the robotic arm

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number |  |  |  |  |
| 0-1 |  |  | 0 | 90 |
| 1-2 |  | 0 |  | 0 |
| 2-3 | -90 | 0 |  | 0 |
| 3-4 |  | 0 | 0 | 90 |
| 4-5 |  |  | 0 | 0 |

represents a rotation about the local z-axis, in another word, is the joint angle. The joint offset which is the distance on z-axis between two successive common normals is defined as . The length of each link is . The twist angle means the angle between two successive z-axes. Notice that the rotation about is . This is because when is 0, there is a angle between link 1 and link 2.

The transformation between two successive frames is expressed by:

()

Where means , means . According to equation (3.42) and Table 4, the transformation of between two joints are:

() ()

() ()

()

So the total transformation between the refernce frame and the end-effector wll be:

()

The aim of this study is that the robotic arm is able to pick up, move and place objects. The objects are at the same horizontal level with the base of the robotic arm. The position and the distance between the objects and the robotic arm can be detected by ultrasonic sensor or camera, so the microcontroller and computer need to use the information to calculate the rotation angle for each motor to make the gripper move to the desired location and pick up objects. Thus, the inverse kinematics of the robotic arm also involves the path planning which tells the robotic arm which position to reach the objects.

There are three paths have been designed to take an object at the ground level. The first path is designed to make the robotic arm reach an object at far territory, so the end-effector must be horizontal. Another path is to reach the objects in the middle range, so the orientation of the gripper keeps at . The wrist angle is fixed in the third path to make end-effector arrive at the position where is closed to the robotic arm. The illustration of the three positions to take objects at three regions shown in Fig. 4.

The distance between the desired objects and the robotic arm can be detected by the ultrasonic sensor. After applying the kinematic model of the robotic arm and path planning, the inverse kinematic equations are able to be calculated. Then, the rotation angles for shoulder, elbow, wrist joints also can be calculated by Arduino.

In order to achieve picking and placing objects in a random position in robotic arm's operation range, the base of the robotic will turn from and an ultrasonic sensor will scan the area. When the sensor detects the object, it will give the distance value between the sensor and objects to the microcontroller. Arduino will calculate the distance and convert to rotation angles for each motor to pick that object, then it will pass commands to the motor driver and motors spin until the robotic arm at the right position and pick up the object. In the end, the robotic arm takes the object to the desired place.

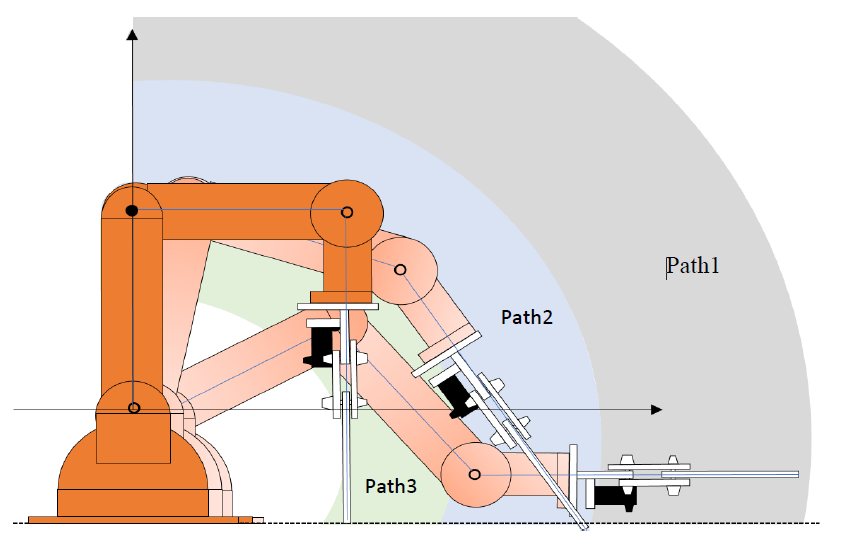


Fig. 4 The three picking up positions of the robotic arm

# computer vision processing

Since the distance sensors send out the ultrasonic wave in a cone, the sensor cannot precisely detect the location of an object or find the centre of an object. So the Picamera can be used for compensation of the disadvantage of the distance sensor to find the precise location of an object. In another word, the camera is used to detect an object and tell the robotic arm to face the object straight and the gripper points at the centre of the object in order to improve the accuracy of the pickup movement.

In order to detect and track an object by a camera, the computer vision technique is involved in the project. The single-board computer Raspberry Pi is used to deal with real-time computer vision. Python programming language and OpenCV (open source computer vision library) have been applied in the computer vision programming in the study.

OpenCV is a programming function library for dealing with image processing and real-time computer vision[5]. The functionality of the library includes video/image input and output, processing, display and facial recognition, object identification. It also includes the machine learning library which contains artificial neural networks, decision tree learning, deep neural networks.

When the camera captures real-time videos, the robotic arm can recognise the objects which it needs to pick up. It must distinguish the objects from background in the video. Because the robotic arm only takes cubic objects in this project, the objects become square or rectangle shape in 2-dimension videos. Thus, shape detection technique is used in programming for object detection.

The robotic arm need face the object straight and the gripper must point at the centre of the object when it picks up the object. In addition, the camera is installed in the middle of the gripper. So the centre of the object should be in the middle of x-coordinate the video. The centre of the objects can be calculated by using the image moment. The Fig. 5 shows the results of object detection and the location of the centre of the objects.



Fig. 5 The object detection and the centre of the objects in the video

# voice control

In this project, the robotic arm can be controlled by voice commands. The voice commands are collected by Amazon Echo and then recognised and analysed using Alexa Voice Service. After voice recognition, the voice commands are transferred to plain text commands to a public web service. So, the Raspberry Pi does not need to be on the same local network with Amazon Echo to access the web service and gets the text commands. Then, the Raspberry Pi gives commands to Arduino by changing its pin status. After Arduino successfully reads the pin status from Raspberry Pi, it will control the robotic arm to do a specific movement. The schematic of the voice control of the robotic arm is in Fig. 6.

A picture containing electronics, circuit

Description generated with high confidence

Fig. 6 The schematic of the voice control of the robotic arm

The public web service is created by ngrok which creates a secure connection from a public endpoint to a locally running web service. It allows the Raspberry Pi to collect the text commands from Alexa Voice Service through any network.

There are three voice commands have been designed for controlling the movements of the robotic arm. The TABLE III shows the movements corresponding the voice command.

TABLE III The movements of the robotic arm controlled by voice command

|  |  |
| --- | --- |
| Voice command | Movement |
| Robotic arm to scan | The basement of the robotic arm starts to rotate from to |
| Robotic arm to pick up an object | The robotic arm finds an object and picks it up, then place the object at a specific place. |
| To stop | The robotic arm stops the present movement and goes back to start position. |

# artificial neural network

As the fast development of artificial intelligence technology, robot control has benefited greatly from it[6]. It has many advantages in robot performance such as machine learning, less computing time and speech recognition. The artificial neural network is a software simulation that is to simulate lots of interconnected neurons inside the computer, which is similar to the biological brain[7]. A large number of neurons are interconnected for information processing to solve a specific problem. The most important feature of the neural network is that it is able to learn by itself through examples, just like a human brain[8].

In this project, supervised learning strategy has been applied in training the neural network, because the correct answers for angle values for the joints in the robotic arm can be calculated by inverse kinematic equations when the distance between objects and the robotic arm are known. Numbers of sets of distance and shoulder, elbow, wrist joints angles of the robotic arm will be used for training the neural network. Thus, the distance between objects and the robotic arm is the input for the neural network and the shoulder, elbow, wrist joints rotation angles of the robotic arm are the three outputs of the neural network.

Neural Network Toolbox in Matlab software has been applied for simulating the artificial neural network. It provides algorithms and pretrained models to create and train the neural network. In the project, there are 20 neurons has been used in the hidden layer of the network. It has been mentioned that there are three different distance regions where the end-effector of the robotic arm can reach. The robotic arm works path 1, path 2 and path 3 in configurations to pick up objects in the three regions and different inverse kinematics equations. Thus, the neural network has used three different sets of value for the training process. There are 100 distance values has been created randomly in each region. the rotation angles of shoulder, elbow and wrist can be calculated by the distance values. Hence, the distance value is the input for the artificial neural network and three angle values are the output. The 300 sets of example values have been applied for training the neural network.

After 4 times training the artificial neural network, the performance becomes more accurate. There are 11 random distance data feeds in the input layer of the neural network to test its performance. The Fig. 7 provides the test results.

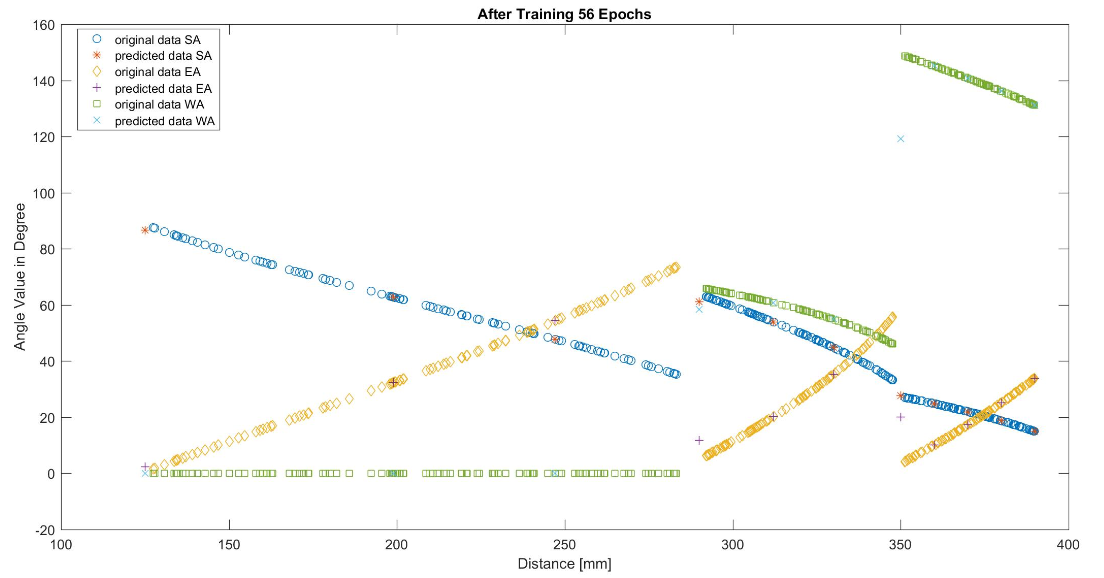


Fig. 7 The results getting from the neural network using 11 test data

In the figure, SA, EA and WA are shoulder angle, elbow angle and wrist angle respectively. According to Fig. 7, most of the angle values which are calculated by the neural network are fallen on the original data. That means the angle values are almost the same with the test data.

The training performance is shown in Fig. 8. The neural network can achieve the Mean Square Error (MSE) performance of 0.028362 at 50 of epochs.

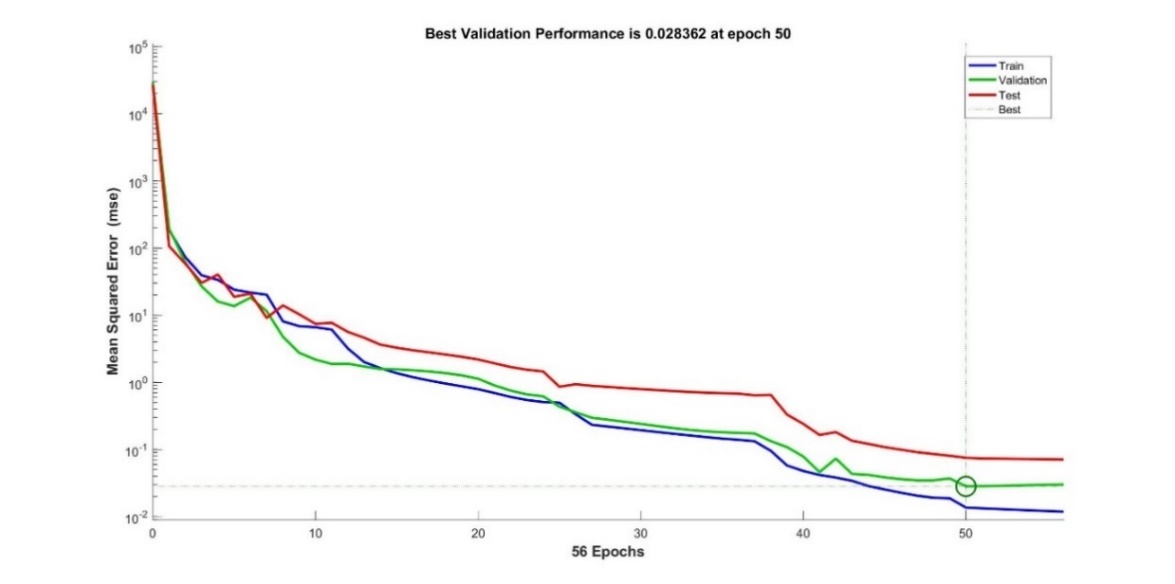


Fig. 8 Training performance results

At the beginning of the training, the error between the correct answer and the output of the neural network is significant because of random initial weight in the network. However, the network adjusts the weights by itself during the training and square error is sharply decreased. After 50 epochs, the network becomes stable and remains the mean square error level. It means the neural network is reliable.

The error histogram result shows most of the error occurs at the range to . That means the error of the neural network is less than . It is very accurate.



Fig. 9 The regression results of training, test and validation

Fig. 9 shows regression data of the training, test and validation. The value R is close to 1 means outputs values of the neural network are approaching the outputs' value of the actual data. In the figure, the R values in both training and test are very close to 1. According to the performance of the training and testing, the artificial neural network has excellent accuracy and can produce the correct joint rotation values for robotic arm by feeding in distance value between objects and the robotic arm.

# Comparison with other methods

A research team from both Princeton University and Massachusetts Institute of Technology has presented a system which is capable of grasping and recognizing both known and novel objects by using deep convolutional neural networks (ConvNets)[9]. There are four pick up behaviors have been developed and the system uses ConvNet to predict affordances for a scene. The candidate objects are recognized by cross-domain image matching and the grasping behaviors adapt to novel objects without additional re-training. The system has 75% pick success grasping rate and 100% recognition accuracy. However, it is more complex and it requires advanced computing machines. The method presents in this paper is not suit for novel objects. The objects recognition is based on shape detection and poses of the objects depend on centroid of the objects. The method is not suit for the shape unknown objects and the grasping success rate will be reduced if the weight of the object is uneven. There is another method for localization of the objects by placing two 2-D cameras attaching to the ground[10, 11]. This method establishes 3-D position and orientation of the objects and the average error rate in the distance to the target object is about 2mm. The method provide precise results but the system is more complex and requires that the robotic arm and the two cameras must be at specific locations. The systems in this paper is less complex and the camera is located at the end-effector. The robotic arm doesn’t have to be at a certain location. It is still able to pick up and recognize objects even the robotic arm has mobility.

# conclusion

In this paper, the robotic arm can pick and place the objects at the random positions based on Arduino microcontroller, simple sensors and cameras. The ultrasonic sensor detects the objects and provides the distance value for the inverse kinematic equation to calculate the rotation angles for each servo motor. Arduino controls the servo motors’ movement, and the robotic arm moves its gripper to arrive at the desired location. The camera captures the video of the objects and Raspberry Pi processes the video and detects the objects and find the centroid of the objects. The information of the computer vision processing is used to tell the robotic arm whether the objects are at the centre of its gripper, in order to improve the accuracy for the robotic arm to pick up. Raspberry Pi communicates with Arduino microprocessor to control the robotic arm to move to the right position. The pick-up movement is successful and accurate.

The voice commands are captured by Amazon Echo and Amazon Echo streams the human speech to Alexa Voice Service to analyse. It produces the plain text command, and Raspberry Pi gets the command from the cloud service and tells the Arduino which movement should be executed.

In addition, the artificial neural network also can control the movement of the robotic arm. After training, the output of the artificial network is able to provide angle value for every joint in the robotic arm when the distance between objects and the robotic arm is the input of the network. The performance of the neural network is accurate and reliable to control the robotic arm.

##### Acknowledgment

We thank London South Bank University for the financial support of this project.

##### References

[1] J. Leitner, "From Vision to Actions Towards Adaptive and Autonomous Robots," PhD, Faculty of informatics, Universita della svizzera italiana, 2013.

[2] R. S. Hartenberg and J. Denavit, "A kinematic notation for lower pair mechanisms based on matrices," *Journal of applied mechanics,* vol. 77, pp. 215-221, 1955.

[3] S. B. Niku, *Introduction to robotics : analysis, control, applications*: Hoboken, N.J : Wiley, 2nd ed., 2011.

[4] A. Roshanianfard and N. Noguchi, "Development of a 5DOF robotic arm (RAVebots-1) applied to heavy products harvesting," *IFAC-PapersOnLine,* vol. 49, pp. 155-160, 2016/01/01 2016.

[5] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Realtime Computer Vision with OpenCV," *Queue,* vol. 10, pp. 40-56, 2012.

[6] A. R. J. Almusawi, L. C. Dülger, and S. Kapucu, "A New Artificial Neural Network Approach in Solving Inverse Kinematics of Robotic Arm (Denso VP6242)," *Computational Intelligence & Neuroscience,* pp. 1-10, 2016.

[7] C. Woodford. (2017, May 08). *Neural networks*. <http://www.explainthatstuff.com/introduction-to-neural-networks.html>

[8] C. Stergiou and D. Siganos. (May 8). *NEURAL NETWORKS*. https://[www.doc.ic.ac.uk/~nd/surprise\_96/journal/vol4/cs11/report.html](http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html)

[9] A. Zeng, S. Song, K. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Alet, N. C. Dafle, R. Holladay, I. Morena, P. Q. Nair, D. Green, I. Taylor, W. Liu, T. Funkhouser, and A. Rodriguez, "Robotic Pick-and-Place of Novel Objects in Clutter with Multi-Affordance Grasping and Cross-Domain Image Matching," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1-8.

[10] S. Kang, K. Kim, J. Lee, and J. Kim, *Robotic vision system for random bin picking with dual-arm robots* vol. 75, 2016.

[11] C. S. Teodorescu, S. Vandenplas, B. Depraetere, J. Anthonis, A. Steinhauser, and J. Swevers, *A fast pick-and-place prototype robot: design and control*, 2016.