Assistive Robot Manipulator Pose Prediction based on Object Orientation using CNN

Nuryono S Widodo
Dept. of Electrical Engineering
Institut Teknologi Sepuluh November
Surabaya, Indonesia, 60111
Dept. of Electrical Engineering
Universitas Ahmad Dahlan
7022211019@mhs.its.ac.id
nuryono.sw@uad.ac.id

Eko Mulyanto Yuniarno
Dept. of Electrical Engineering
Dept. of Computer Engineering
Institut Teknologi Sepuluh November
Surabaya, Indonesia, 60111
ekomulyanto@ee.its.ac.id

Mauridhi Hery Purnomo
Dept. of Electrical Engineering
Dept. of Computer Engineering
Institut Teknologi Sepuluh November
Surabaya, Indonesia, 60111
hery@ee.its.ac.id

Ronny Mardiyanto
Dept. of Electrical Engineering
Institut Teknologi Sepuluh November
Surabaya, Indonesia, 60111
ronny@elect-eng.its.ac.id

***Abstract*— Robot, as dictated by its assistive function for human, frequently needed to perform object manipulation (i.e. robotic manipulation). A successful robotic manipulation not only determined by the robot actuator, it is also strongly affected by the robot perception in capturing the world and the workpieces. Robotics has made extensive use of camera-based visual perception, especially robotic manipulation. This perception will enable the robot to perform object(workpieces) detection and localization and generate suitable control policy to move the robot via a series of robot poses. This paper proposed an assistive robot manipulator pose prediction based on object orientation using CNN. The manipulator pose prediction will enable further action for a successful robotic manipulation by providing a relative pose between the object and manipulator. Experiment result shown that the presented method was able to predict the robot pose, especially on the fifth joint of the robot accurately based on the object orientation images.**

Keywords—convolutional neural network, assistive robot manipulator, visual perception

I. INTRODUCTION

Humans are naturally capable of varying the interaction and manipulation of various objects in diverse environments, this ability is supported by the existence of a human perception system consisting of at least visual perception and tactile perception as well as manipulators and actuators possessed by humans in the human body. Robot as a device that is expected to help humans is of course also required to have interaction and object manipulation capability even though it is still within various limits. Research related to this includes sorting and retrieving objects [1], as well as on daily activities in the form of interaction or manipulation of room doors [2], [3] or cupboard doors [4].

II. RELATED WORKS

Studies in [5], [6] have attempted to solve challenges related to cluttered scenes and novel objects by utilizing an Artificial Intelligence (AI) based approach. Bousmalis et al.[5] employed convolutional neural network (CNN) approach to teach a grasping system to grasp novel objects from unprocessed monocular RGB images, using simulated environments and domain adaption techniques. In [6] Kumra & Kanan presents a CNN based robotic grasp detection method that uses an RGB-D scene image to predict the ideal parallel gripper grasping pose for novel objects. Whereas [7] solved grasping unknown objects with soft hands problem using a three-dimensional deep convolutional neural network (3D CNN) method. CNN based approach also found in [8]–[11]. Meanwhile [12] proposed a reinforcement learning (RL) technique for manipulating and grasping for a mobile manipulator in order to address changing manipulation dynamics and unpredictable external disturbances. RL type of machine learning based on force and displacement data was also used in [13] to perform stiffness and position control optimization in a soft robot arm module. Another application of RL also found in [14]. AI-based approaches can also be found in studies related to manipulation [15]. In term of sensor or input modalities, most studies used RGB, RGB-D images or depth images, as mentioned in [5]–[12] . The perception that has been used in these studies is primarily a perception based on a vision system or visual perception which is generally based on a camera. The use of camera and CNN approach is quite successful, especially in object detection and object position sensing as well as generating control for object manipulation. Other types of modalities such as force and displacement, tactile can be found in [13], [15].

The study presented in this paper aims to provide an approach to infer robot manipulator pose using machine learning image classification technique.

III. METHOD

This paper proposed a pose prediction system for a digitally controlled 6 DOF manipulator equipped with a gripper and camera that relies on information from a visual sensor. Visual sensor will be used to acquire visual information about the manipulated object. Realization of the proposed system will be validated in experiment's measurements, mainly on the detection success rate. In the software part a CNN based detection implemented.

A. Manipulator hardware structure

The experiment was conducted on a 6 DOF robot manipulator that equipped with a UVC webcam attached to its end effector. Geometrical design of the manipulator presented in Fig.1.

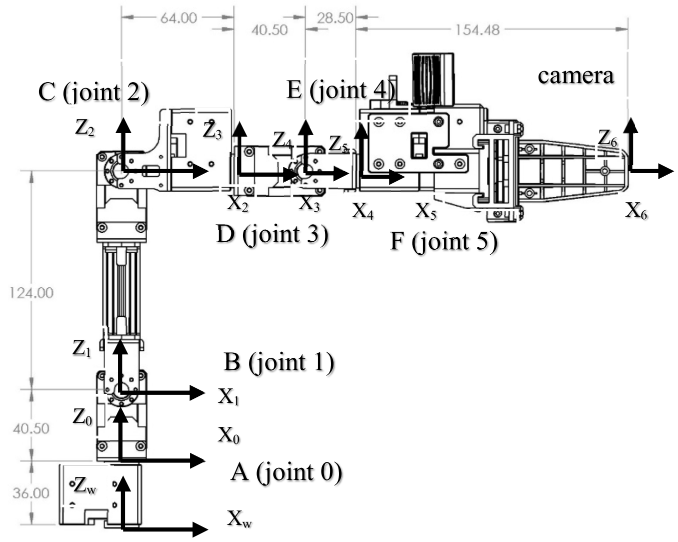


Fig. 1. Gometrical Design of the 6-DOF Manipulator

The Manipulator consisting of six joints starting from joint 0 to joint 5 with link1(L_1), link2(L_2), link3(L_3), link4(L_4), link5(L_5), link6(L_6), link7(L_7) as the links with each link dimensions as follows L_1=36mm, L_2=40.5mm L_3=124mm, L_4=64mm, L_5=40.5mm, L_6=28.5mm, L_7=154.48mm. The manipulator had a parallel griper with 20-75mm openings as its end effector.

Each joint of the manipulator actuated by a servomotor having 0.088° rotation resolution and 4.1Nm torque when powered by a 12VDC power source. Reference Coordinate (the world coordinate system) and each joint coordinate defined and depicted as Fig.2.

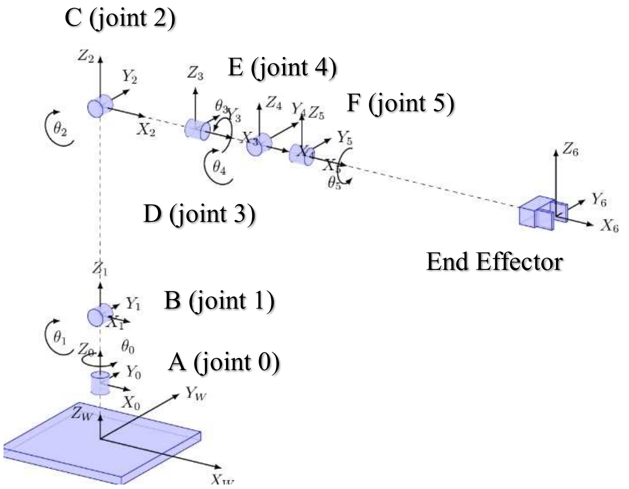


Fig. 2. World & Joint Coordinate System

As refer to Fig.1 and Fig. 2, it can be written that the 6 DOF manipulator has transformation matrix as follows:

For

$$ {}^{W}_{0}T = \begin{bmatrix} \cos\theta_0 & -\sin\theta_0 & 0 & 0 \\ \sin\theta_0 & \cos\theta_0 & 0 & 0 \\ 0 & 0 & 1 & 36 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

$$ {}^{0}_{1}T = \begin{bmatrix} \cos\theta_1 & 0 & \sin\theta_1 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_1 & 0 & \cos\theta_1 & 40.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2) $$

$$ {}^{1}_{2}T = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 & 124 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3) $$

$$ {}^{2}_{3}T = \begin{bmatrix} 1 & 0 & 0 & 64 \\ 0 & \cos\theta_3 & -\sin\theta_3 & 0 \\ 0 & \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4) $$

$$ {}^{3}_{4}T = \begin{bmatrix} \cos\theta_4 & 0 & \sin\theta_4 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_4 & 0 & \cos\theta_4 & 40.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5) $$

$$ {}^{4}_{5}T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_5 & -\sin\theta_5 & 0 \\ 0 & \sin\theta_5 & \cos\theta_5 & 28.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6) $$

$$ {}^{5}_{6}T = \begin{bmatrix} 1 & 0 & 0 & 154.48 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7) $$

Then the coordinate transformation from the world coordinate system to the end effector robot is obtained as follows:

$$ {}^{W}_{6}T = {}^{W}_{0}T\,{}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T\,{}^{3}_{4}T\,{}^{4}_{5}T\,{}^{5}_{6}T $$

where ${}^{m}_{n}T$ and θ_i as transformation matrix between coordinate system m and n, whereas θ_i as the joint angle.

B. Robot Manipulator Pose Prediction based on Object Orientation using CNN

- Robot Manipulator pose prediction based on object orientation image constructed from training data acquired from the end effector camera, with 640x480 resolution comprises of 11 classes from 11 different manipulator pose from fifth joint rotation i.e. joint 4 (θ_4) value variation while keeping all other joints fixed (θ_0 = 0°, θ_1 = - 45°, θ_2 = 85°, θ_3 = 0°, θ_5 = 0°). Image data were acquired by placing a daily activity object on a 300 mm distance from the manipulator base as shown in Fig.3., followed by controlling the manipulator to an initial position, then a sequential 5° rotation commands were sent to the robot. There are short delays between each command, in which the camera capture the object.

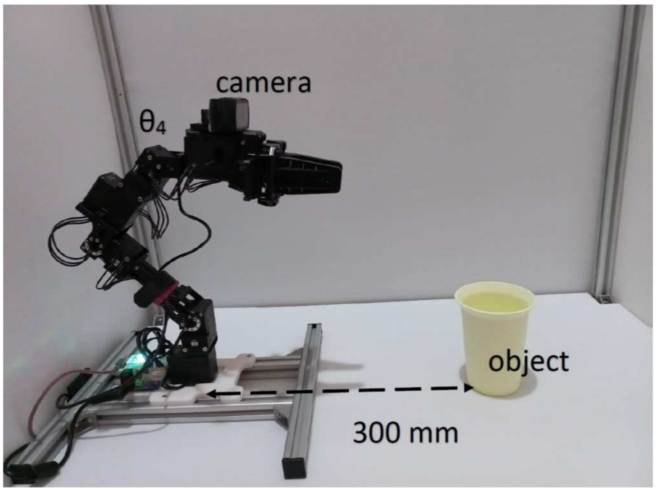
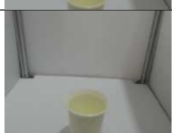


Fig. 3. Proposed Assistive Robot Manipulator

- The θ_4 value and its corresponding object orientation images and presented in Table I below:

TABLE I. OBJECT ORIENTATION IMAGES

No.	Joint value and corresponding images	
	joint 4 (θ_4) value	*Image*
1.	45°	
2.	50°	
3.	55°	
4.	60°	
5.	65°	
6.	70°	
7.	75°	
8.	80°	
9.	85°	
10.	90°	
11.	95°	

- Images acquired from the aforementioned scenario were fed to a Convolutional Neural Network having a model listed in TABLE II.

TABLE II. CNN MODEL

No.	CNN Model		
	Layer (type)	*Output Shape*	*Param #*
1.	input_1 (InputLayer)	[(None, 128, 128, 3)]	0
2.	conv2d (Conv2D)	(None, 128, 128, 32)	896
3.	max_pooling2d (MaxPooling2D)	(None, 64, 64, 32)	0
4.	conv2d_1 (Conv2D)	(None, 64, 64, 32)	9248
5.	max_pooling2d_1 (MaxPooling2D)	(None, 32, 32, 32)	0
6.	conv2d_2	(None, 32, 32, 32)	9248
7.	flatten (Flatten)	(None, 32768)	0
8.	dense (Dense)	(None, 100)	3276900
9.	dense_1 (Dense	(None, 100)	10100
10.	dense_2 (Dense)	(None, 11)	1111
Total params: 3,307,503 Trainable params: 3,307,503 Non-trainable params: 0			

- The CNN take an input size of 128x128, from a resized and normalized 640x480 image. The first convolution layer, a 2D convolution layer having 32 kernels with 3x3 in size, and ReLU activation function, the padding resulting a same size output. Max pooling layer with 2x2 filter used then resulting 64x64 output. The next convolution layer has the same number and size of kernel as the first convolution layer. The next max pooling layer has 2x2 filter followed by a 2D convolution layer with 32 kernel and 3x3 in size and

ReLU activation. A Flatten layer followed by 3 dense layers resulting 11 outputs.

- Training was conducted in 100 epochs using 110 images with the result illustrated in Fig.4.

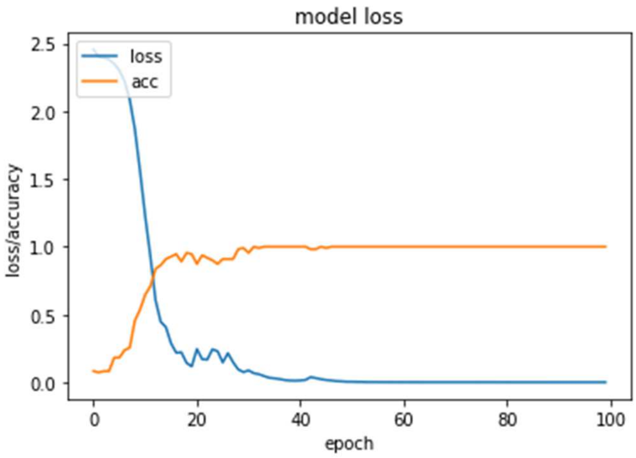


Fig. 4. Training Result

IV. Result & Discussion

The training result graphic shows that the training accuracy reach around 1.0 after 20 epochs, and became stabilized after 40 epochs. The loss value decreases rapidly from the beginning of training and stabilized after 50 epochs.

The model was tested to perform classification of 44 different images. The classification result shown in Table III below

TABLE III. Classification Result

Actual	Predicted joint value										
	45	*50*	*55*	*60*	*65*	*70*	*75*	*80*	*85*	*90*	*95*
45	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
55	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
60	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
65	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
75	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.25	0.00
90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Based on the classification result it can be concluded that the proposed system is able to predict change of robot manipulator pose based on object orientation image captured from the attached camera.

Acknowledgment

We would like to express our sincere gratitude to all members of Visicom Laboratory ITS for the assistances in preparing this manuscript. The authors also grateful to the UAD Laboratory for the equipment provided to conduct this research.

References

[1] R. Krug *et al.*, "The Next Step in Robot Commissioning: Autonomous Picking and Palletizing," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 546–553, 2016, doi: 10.1109/LRA.2016.2519944.

[2] B. Axelrod and W. H. Huang, "Autonomous door opening and traversal," in *IEEE Conference on Technologies for Practical Robot Applications, TePRA*, 2015, vol. 2015-Augus, pp. 1–6, doi: 10.1109/TePRA.2015.7219686.

[3] H. Tsukagoshi, M. Watanabe, T. Hamada, D. Ashlih, and R. Iizuka, "Aerial manipulator with perching and door-opening capability," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2015, vol. 2015-June, no. June, pp. 4663–4668, doi: 10.1109/ICRA.2015.7139845.

[4] Y. Karayiannidis, C. Smith, F. E. V. Barrientos, P. Ogren, and D. Kragic, "An Adaptive Control Approach for Opening Doors and Drawers under Uncertainties," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 161–175, 2016, doi: 10.1109/TRO.2015.2506154.

[5] K. Bousmalis *et al.*, "Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 4243–4250, 2018, doi: 10.1109/ICRA.2018.8460875.

[6] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 2017-Septe, pp. 769–776, 2017, doi: 10.1109/IROS.2017.8202237.

[7] C. Choi, W. Schwarting, J. Delpreto, and D. Rus, "Learning Object Grasping for Soft Robot Hands," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2370–2377, 2018, doi: 10.1109/LRA.2018.2810544.

[8] J. Lundell, F. Verdoja, and V. Kyrki, "DDGC: Generative Deep Dexterous Grasping in Clutter," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6899–6906, 2021, doi: 10.1109/LRA.2021.3096239.

[9] H. Zhang, Y. Wu, E. Demeester, and K. Kellens, "BIG-Net: Deep Learning for Grasping with a Bio-Inspired Soft Gripper," *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 584–591, 2023, doi: 10.1109/LRA.2022.3229237.

[10] Y. Shi, Z. Tang, X. Cai, H. Zhang, D. Hu, and X. Xu, "SymmetryGrasp: Symmetry-Aware Antipodal Grasp Detection From Single-View RGB-D Images," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 12235–12242, 2022, doi: 10.1109/LRA.2022.3214785.

[11] V. Satish, J. Mahler, and K. Goldberg, "On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1357–1364, 2019, doi: 10.1109/LRA.2019.2895878.

[12] Z. Li, T. Zhao, F. Chen, Y. Hu, C. Y. Su, and T. Fukuda, "Reinforcement Learning of Manipulation and Grasping Using Dynamical Movement Primitives for a Humanoidlike Mobile Manipulator," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 121–131, 2018, doi: 10.1109/TMECH.2017.2717461.

[13] Y. Ansari, M. Manti, E. Falotico, M. Cianchetti, and C. Laschi, "Multiobjective Optimization for Stiffness and Position Control in a Soft Robot Arm Module," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 108–115, 2018, doi: 10.1109/LRA.2017.2734247.

[14] T. G. Thuruthel, E. Falotico, F. Renda, and C. Laschi, "Model-Based Reinforcement Learning for Closed-Loop Dynamic Control of Soft Robotic Manipulators," *IEEE Trans. Robot.*, vol. 35, no. 1, pp. 127–134, 2019, doi: 10.1109/TRO.2018.2878318.

[15] S. Funabashi, A. Schmitz, T. Sato, S. Somlor, and S. Sugano, "Versatile In-Hand Manipulation of Objects with Different Sizes and Shapes Using Neural Networks," *IEEE-RAS Int. Conf. Humanoid Robot.*, vol. 2018-Novem, pp. 768–775, 2019, doi: 10.1109/HUMANOIDS.2018.8624961.