American Journal of Mechanical Engineering, 2014, Vol. 2, No. 7, 216-218
Available online at http://pubs.sciepub.com/ajme/2/7/9
© Science and Education Publishing
DOI:10.12691/ajme-2-7-9

Robotic Grasping System Using Convolutional Neural Networks

Pavol Bezák[1], Yury Rafailovich Nikitin[1], Pavol Božek[2,*]

[1]Institute of Applied Informatics, Automation and Mathematics, Faculty of Materials Science and Technology,
Slovak University of Technology, Trnava, Slovakia
[2]Kalashnikov Izhevsk State Technical University, Mechatronic Systems Department, Izhevsk, Russia
*Corresponding author: pavol.bozek@stuba.sk

Received September 15, 2014; Revised September 29, 2014; Accepted October 15, 2014

Abstract Object grasping by robot hands is challenging due to the hand and object modeling uncertainties, unknown contact type and object stiffness properties. To overcome these challenges, the essential purpose is to achieve the mathematical model of the robot hand, model the object and the contact between the object and the hand. In this paper, an intelligent hand-object contact model is developed for a coupled system assuming that the object properties are known. The control is simulated in the Matlab Simulink/ SimMechanics, Neural Network Toolbox and Computer Vision System Toolbox..

Keywords: *robot hand, modeling, grasping, convolutional neural networks, deep learning, object recognition, pose estimation*

Cite This Article: Pavol Bezák, Yury Rafailovich Nikitin, and Pavol Božek, "Robotic Grasping System Using Convolutional Neural Networks." *American Journal of Mechanical Engineering*, vol. 2, no. 7 (2014): 216-218. doi: 10.12691/ajme-2-7-9.

1. Introduction

Robotics is moving towards the research and development of technologies that allow the introduction of robots in our daily life. The optimal robot assistant should share a human environment and be able to cope with human presence and interact in a very friendly way. A number of problems need to be solved to create such applications, including transposing the movements used in everyday tasks, as well as finding out how to interpret human interactions and how to use all this knowledge to create robots that can successfully act as assistants. The need of having intelligent robots means that the complexity of programming must be greatly reduced, and robot autonomy must become much more natural. This challenge is relevant to a new generation of robots, which must interact with people, and operate in human environments [1,2,3].

The principal task in grasping involves the interaction between the object and the hand. Let's consider a fixed object with known coordinates and a finger of the robot hand. The task of grasping can be divided into two distinct problems. First task is to command the hand to reach the object location. Contact force is generated when the hand hits the object, which needs to be dealt with. The contact force needs to be bounded to prevent any slippage or damage of the object [4,5,6].

2. CAD Model of Simplified Humanoid Robot Hand

After kinematic analysis the 3D model of simplified (three-fingered) humanoid hand can be created. For this task the software Autodesk Inventor was used. In the CAD model we did not assume actuators. The CAD model (Figure 3) serves us only for better imagination and for observation of possible motions and operations with hand [7,8].

It is possible to import the model from Autodesk Inventor to Matlab Simmechanics to further analysis. We used the tool smlink_linkinv. After successful import we gained also Simulink model of the model of the hand.

In the Matlab Simulink model we can see parts of the simulated robotic hand after import from Autodesk Inventor (Figure 1). The model had to be edited to add the required functionality.

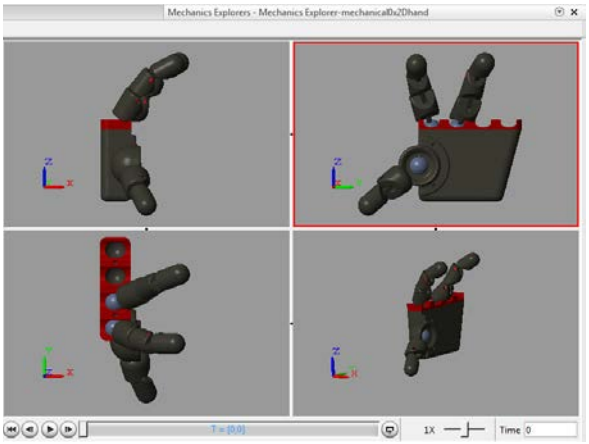

Figure 1. Imported CAD model of humanoid robot hand in Matlab Mechanics Explorer

3. Object Recognition and Pose Estimation

3.1. Brief Introduction to Deep Neural Networks and Deep Learning

In machine learning, a deep belief network (DBN) is a generative graphical model, or alternatively a type of deep neural network, composed of multiple layers of latent variables ("hidden units"), with connections between the layers but not between units within each layer [9,10]. Deep learning is a set of algorithms in machine learning that attempt to model high-level abstractions in data by using model architectures composed of multiple non-linear transformations [11]. Various deep learning architectures such as deep neural networks, convolutional deep neural networks, and deep belief networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, and music/audio signal recognition where they have been shown to produce state-of-the-art results on various tasks [12].

3.2. Convolutional Neural Networks (CNN)

A CNN is composed of one or more convolutional layers with fully connected layers (matching those in typical artificial neural networks) on top. It also uses tied weights and pooling layers. This architecture allows CNNs to take advantage of the 2D structure of input data. In comparison with other deep architectures, convolutional neural networks are starting to show superior results in both image and speech applications. They can also be trained with standard back propagation. CNNs are easier to train than other regular, deep, feed-forward neural networks and have many fewer parameters to estimate, making them a highly attractive architecture to use.

Comparing with the traditional object recognition based on the deep learning model, we focus on the object pose estimation including object recognition. Deep learning methods have the capability of recognizing or predicting large set of patterns by learning sparse features of small set of patterns. With this advantage, we can use a small set of poses to train the deep learning model, and then predict a large set of poses with the model.

For general objection recognition and image classification tasks, variants of Convolutional Neural Networks (CNNs) have emerged as robust supervised feature learning and classification tools, especially when combined with max-pooling (MPCNN) (Figure 2) [13].

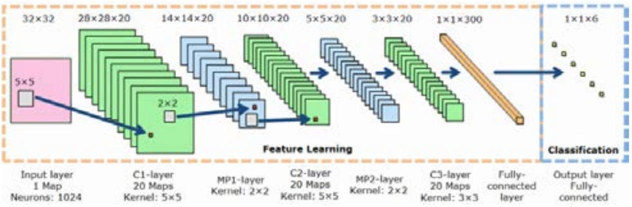

Figure 2. MPCNN architecture using alternating convolutional and max-pooling layers [13]

MPCNNs include convolutional layers and subsampling layers. MPCNNs are different according to the variety of training and realization of convolutional and subsampling layers.

3.3. Convolutional Layer

The parameters of the convolutional layer are: the number of maps, the size of the maps and kernel sizes. Each layer (L) includes maps (M). A kernel (K) of size is shifted over the valid region of the input image. Each map in Layer Ln is connected to all maps in layer Ln-1. Neurons of a given map share their weights but have different input fields [13].

3.4. Max-pooling Layer

The output of the max-pooling layer is determined by the maximum activation over non-overlapping rectangular regions. Max-pooling improves generalization performance [13].

3.5. Classification Layer

To complete the MPCNN, a shallow Multi-layer Perceptron (MLP) is used. The output layer has one neuron per class in the classification task [13].

4. Experiment

4.1. Object Detection

The system consists of the Matlab SimMechanics model of robotic hand scene with objects and simulated camera. The virtual objects are in the reach of the vision system and are recognized through the Matlab Computer Vision Toolbox with implemented model of MPCNN.

The input images come from RGBD camera data that opposed to simple 2D image data has been shown to significantly improve the grasp detection results.

Visual object detection is the first key action for robotic grasping. It is needed to use or develop reliable methods that effectively recognize foreground objects that are present at the input to the vision system and detect the objects from the background.

One of the possible approaches is to use sparse coding or K-means clustering. The first step is to build dictionary of objects. Clustering enables us to separate groups of color components of the images with background and objects. Each component has a location in feature space and it is important to find partitions such that components within each cluster are as close to each other as possible and as far from components in other clusters as possible [14].

4.2. Object Recognition and Pose Estimation

Using the simulation, we implemented object pose estimation and pose estimation using the methods of deep learning, which have the capability of recognizing or predicting large set of patterns by learning sparse features of small set of patterns. With this advantage, we can use a small set of poses to train the deep learning model, and then predict a large set of poses with the model [15].

4.3. Robotic Graspin

This stage presents the system that changes the position and orientation of the gripper in order to grasp the objects.

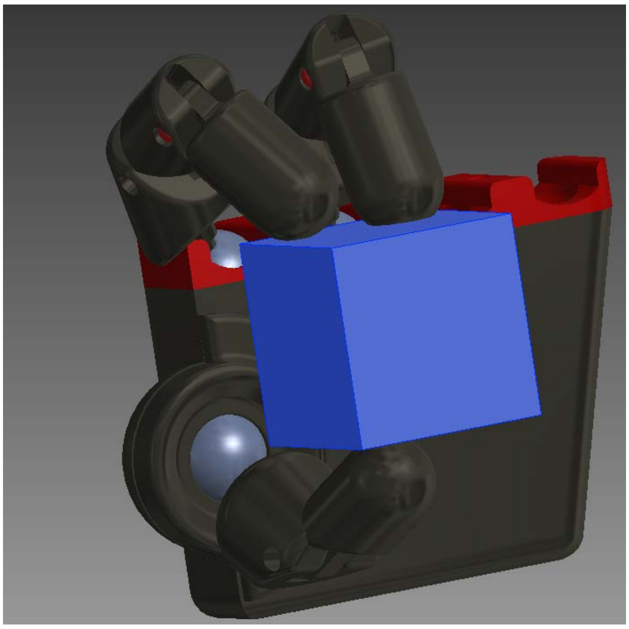

Figure 3. 3D model of humanoid robot hand grasping the object

4.4. Simulation Results

The developed models described in previous sections have been implemented in selected Matlab Toolboxes (Computer Vision System Toolbox, Deep Learning Toolbox and SimMechanics).

5. Conclusion

This paper represents a simulated model of a multi-fingered robotic hand for grasping tasks. The model includes kinematics, dynamics, object representation and contact modelling [16,17]. The contact model determinates the forces that are applied to the object by the robotic hand. The object detection, object recognition and robotic hand pose estimation are based on Max-pooling Convolutional Neural Networks – one of the most popular deep learning models. Using deep learning enables to avoid hand-engineering features, learning them instead.

Acknowledgement

The study on the subject “Development of intelligent control and diagnostic systems of mechatronic drives” has been made with financial support from the Ministry of Education and Science of the Russian Federation within the framework of the governmental task.

References

[1] Collet, A., Martinez M., Srinivasa, S. S., *The MOPED framework: Object Recognition and Pose Estimation for Manipulation*, in International Journal of Robotics Research, 2011.

[2] Hasan, R.; Rahideh, A; Shaheed, H., *Modeling and interactional control of the multifingered hand*, 19th International Conference on Automation and Computing (ICAC), pp. 6-10, 13-14 Sept. 2013.

[3] Pellerin, Ch.l, *The Salisbury Hand*, Industrial Robot: An International Journal, Vol. 18(1991) Issue: 4, pp. 25-26.

[4] Townsend, W.T., *The Barrett Hand grasper-programmably flexible part handling and assembly*, Industrial Robot: An International Journal, Vol. 10(3), 2000, pp. 181-188.

[5] Grebenstein, M., *The DLR hand arm system*, IEEE International Conference on Robotics and Automation (ICRA), 2011, pp. 3175-3182.

[6] Jacobsen, S., *Design of the Utah/M.I.T. dexterous hand*, IEEE International Conference on Robotics and Automation (ICRA), Vol. 3, 1986, pp. 1520-1532.

[7] Corrales, J.A; Jara, C.A; Torres, F., *Modelling and simulation of a multi-fingered robotic hand for grasping tasks*, 11th International Conference on Control Automation Robotics & Vision (ICCARV), vol., no., pp. 1577-1582, 2010.

[8] Sang-Mun Lee; Kyoung-Don Lee; Heung-Ki Min; Tae-Sung Noh; Jeong-Woo Lee, *Kinematics of the Robomec robot hand with planar and spherical four bar linkages for power grasping*, IEEE International Conference on Automation Science and Engineering (CASE), pp. 1120-1125, 2012.

[9] Akdagli, A., Toktas, A., Kayabasi, A., Develi, I., *An application of artificial neural network to compute the resonant frequency of e-shaped compact microstrip antennas*, Journal of Electrical Engineering, Vol. 64, No. 5, 2013, 317-322

[10] Ficuciello, F.; Villani, L., *Compliant hand-arm control with soft fingers and force sensing for human-robot interaction*, 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 1961-1966, 2012.

[11] Jincheng Yu; Kaijian Weng; Guoyuan Liang; Guanghan Xie, *A vision-based robotic grasping system using deep learning for 3D object recognition and pose estimation*, IEEE International Conference on Robotics and Biomimetics (ROBIO), pp.1175-1180, 2013.

[12] Karavaev, Y., Klekovki, A., Božek, P., *The implementation of microprocessor device for drilling process monitoring based on artificial neural network*, International Conference on Process Control, pp.163-167, 2013.

[13] Nagi, J.; Ducatelle, F.; Di Caro, G.A; Ciresan, D.; Meier, U.; Giusti, A; Nagi, F.; Schmidhuber, J.; Gambardella, L.M., *Max-pooling convolutional neural networks for vision-based hand gesture recognition*, IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp.342-347, 2011.

[14] Lenz, I.,, Lee, H., Saxena, A., *Deep learning for detecting robotic grasps*, in Proc. Of Robotics: Science and Systems (RSS), 2013.

[15] Mekk, H., Chtourou, M., *Variable structure neural networks for adaptive robust control using evolutionary artificial potential fields*, Journal of Electrical Engineering, Vol. 64, No. 1, 2013, 3-11

[16] Frankovský, P., Hroncová, D., Delyová, I., Virgala, I, *Modeling of Dynamic Systems in Simulation Environment MATLABSimulink-SimMechanics*, American Journal of Mechanical Engineering. Vol. 1, no. 7 (2013), p. 282-288.

[17] Virgala, I., Frankovský, P., *Locomotion mechanism for pipe inspection tasks*, Ad Alta. Vol. 3, no. 2 (2013), p. 84-86.