The Shopping Assistant Robot design based on ROS and Deep Learning

Hang Su, Yusi Zhang, Jingsong Li, Jie Hu
Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education
Wuhan University of Technology
Wuhan, China
hangsu@whut.edu.c

***Abstract*—As the traditional service robots' artificial intelligence bottlenecks, there is a huge gap between service robots and human intelligence in the cognitive and learning discipline. So the service robots could not be widely applied. As deep learning theory is proposed in 2012, it may lead to a generation leap forward in machine learning discipline, so as to improving the traditional robot's cognitive algorithms. This paper studies the principle of three-dimensional Kinect sensor and the deep learning framework of CNN. We proposed a shopping assistant robot designing method which combined Robot Operation System and deep learning method. Firstly, the ROS packages for the service robot are designed. Secondly, Kinect sensors are used for acquiring the information in the robot. Finally, we use simulation to evaluated the design.**

Keywords—Deep Learning, Robot Operation System, Kinect, Robot.

I. INTRODUCTION

With the integration between the information technology and industrialization, the intelligent industry is booming especially as the representative of robotics technology. It becomes an important innovation symbol of the modern science and technology. When Google's AlphaGo beat Lee Sedol in a five-game match, it is a great step of the machine intelligence to the territory of human intelligence [1]. The new intelligence era is coming as the sign of the leap forward from internet plus to the smart plus era.

It is 60 years after the birth of artificial intelligence in 2016. The artificial intelligence development has gone through three stages: (1) The artificial intelligence has achieved rapid development after the Dartmouth meeting in 1956. 30 mathematical principles in about 50 could be proved by machines in 1970s. The scientists felt that all the principles will be able to prove in next 10 years, but it is not the reality. (2) In the 1990s, with the emergence of the Hopfield network and BP (back propagation) algorithm, artificial intelligence had re-emerged. For example, Japan ambitiously proposed the fifth generation of computer artificial intelligence program which later appeared stagnantly. (3) As the current outbreak wave of artificial intelligence, the technology includes not only the depth of the neural network, but also the big data from cloud computing and mobile Internet. The outbreak of this round of artificial intelligence is a real industry outbreak. It is able to change millions of households [2].

Due to the bottleneck of artificial intelligence, the service robot has a huge gap in the field of cognition and learning with the existing human intelligence. So many service positions could not be replaced by robots. The theory of deep learning studied from 2012 may lead to the leapfrogging of machine learning, which may replace the traditional robot cognitive algorithms.

A shopping assistant robot design is presented by our research group in this paper. Firstly, we study the algorithms in deep learning. Secondly, we design a system using ROS, cloud server and Kinect sensor. Finally, we imply the system on the turtlebot2 hardware.

II. RELATED WORK

A. ROS

ROS (Robot Operating System) is a software platform for robot, which provides similar functions as operating system for heterogeneous computer cluster [3].

There are three main features in ROS: distributed computing, portability, ease of testing. Robot with ROS could implement distributed computing, and it is easy to achieve multi-machine communication, improve operational efficiency and real-time operation of the robot. Furthermore, it is easy to reduce duplication of work that people could use and share ROS feature packages which have been finished by many of our predecessors. When you are debugging, you could play back to test many functions by recording the sensor data at one time.

ROS is mainly based on Ubuntu operating system currently. There are three levels in ROS: File system level, Computation Graph level and Community level.

By 2015, Massachusetts Institute of Technology, Stanford University and Carnegie Mellon University funded by Defense Advanced Research Projects Agency (DARPA) have conducted some research about humanoid robot environmental scanning, driving, autonomous action research under hazardous disposal and complex environment.

On March 2016, Boston Dynamics company which belongs to Google demonstrated an automatic stacking shelves robot whose name was ATLAS. The robot could pick up and drop off 10 pounds of goods automatically, as shown in Fig.1.

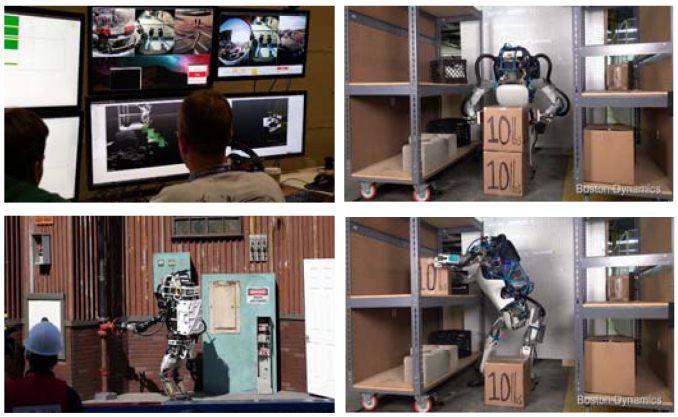

Fig. 1. The Robots design by Stanford, Carnegie Mellon university and Google.

B. *Kinect Sensor*

Microsoft company released a RGB-D sensor, called Kinect V 1.0 in 2010. It released updated version 2 on May 2013, named Kinect V2.0, which is shown in Fig.2. Kinect Sensor is a kind of device shaped like a network camera. It has three lenses: The lens on the left is a RGB color camera with a resolution of 1080P, the lens on the right are respectively infrared transmitter and infrared CMOS camera which constitutes the 3D structure optical depth sensor. There is a built-in microphone array (four sets of microphone) in the lower part of Kinect. The noise will be eliminated after comparing the input of the four sets of microphone at the same time.

TABLE I. KEY FEATURES OF KINECT V2 SENSOR

Key feature	Parameters
color camera	1920 x 1080 pixels 30 Hz (15 Hz in low light)
Infrared (IR) camera	512 x 424 pixels 30Hz
Field of view	70 x 60 degrees
Operative measuring range	From 0.5 to 4.5 meters
Multi-array microphone	4 microphones
Object pixel size (GSD)	Between 1.4mm (@0.5 m range) and 12 mm (@4.5 m range)
Body tracking	Six complete skeletons, 25 joints per person
Sensor dimensions (Length x width x height)	22.9 x 6.6 x 6.7 cm
Weight	Approximately 1.4 kg

There are lots of improvements in the Kinect V2 than Kinect V1, as shown in Table 1. The color camera captures full 1080p video that can be displayed in the same resolution as the viewing screen. In addition to allowing the sensor to be functional in the dark, the IR capabilities produce a lighting-independent view, and the IR and color camera could be used simultaneously. With higher depth fidelity and a significantly improved noise reduction, the sensor could provide improved 3D visualization, improved ability to observing smaller objects. It also improves the stability of body tracking.

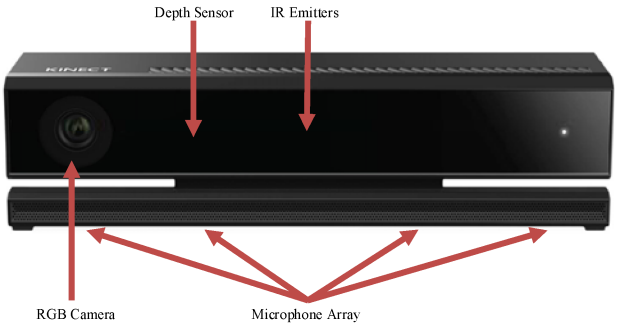

Fig. 2. The Structure of Kinect V2

III. DEEP LEARNING

A. *Deep Belief Networks*

Deep Belief Networks (DBN) is a probability generation model, which contains multiple hidden layer units. It is a composite model which combines multi-layer simple learning model. The deep belief network uses the pre-training result of the Deep Neural Network (DNN) as the initial weight of the network, and then uses the back propagation algorithm or other algorithms to fine-tune the weights. It is very useful for the lack of training data, because the initial weight will have a significant impact on the final performance of the model, and pre-training weights than random weights closer to the optimal Weight. This method could not only improve the performance of the model, but also speed up the convergence speed when tuning.

Deep belief network is composed of multi-layer Restricted Boltzmann Machine (RBM), which could be trained by effective unsupervised greedy training method. Restricted Boltzmann Machine is an energy-based probability generation model, including a visual layer and a hidden layer. There is connection only between the visual layer and hidden layer, and there is no connection between the visual layer units and hidden layer units. The training method for single-layer RBM was originally proposed by Geoffrey Hinton in the training of expert product, known as the contrast divergence (CD) algorithm. An approximation to the maximum likelihood estimate is proposed for the divergence algorithm and is used in the learning of RBM weights. Although the approximation to the maximum likelihood estimate for the divergence algorithm is very crude, since the divergence has not been followed by any gradient of the function, empirical results show that the algorithm is very effective for training depth structures.

RBM training process and the optimization of the weights are as follows:

1) The weights matrix w, the visual layer bias vector b and the hidden layer bias vector a are randomly initialized to the network.

2) The original input data are assigned to the visual layer unit, the visual layer input matrix v is propagated forwardly, the activation probability p (h | v) of the hidden layer unit's output matrix h is calculated by the visual layer, and the input matrix the probability of the forward propagation is obtained by the matrix of the activation probabilities corresponding to the product of v and h.

3) The output of p (h | v) in step b is the probability value of h, which is randomly binarized into a binary variable.

4) Using the probability value of the binarized h in step 3 to propagate in the opposite direction, calculate the activation probability of the matrix v' of the visual layer, and get a reconstruction of the visual layer.

5) Further propagate v' to calculate the activation probability of the hidden layer h'. As in step a, the matrix of the activation probabilities of the input matrices v' and h' corresponds to the probability of back propagation.

6) The activation probability of h' obtained in step 5 is subtracted from the activation probability of the hidden layer h obtained in step 2. The result of the activation probability of the visual layer v subtracting the activation probability of v' is the offset a corresponding to the visual layer v. Subtract the back-propagation probability vector obtained in step 5 from the forward propagation probability vector obtained in step 2. The obtained result is the weight increment between the input layer and the output layer. In each iteration, the updating of the weights and the updating of the biases are performed simultaneously, so they should converge at the same time.

7) Repeat steps 2 through f until convergence or maximum number of iterations is reached. Thus, an RBM training is completed.

B. Convolutional Neuron Networks

Convolutional Neuron Networks (CNN) consist of one or more convoluted layers and a fully connected layer on the top, and include correlation weights and pooling layers. This structure allows the CNN to make use of the input data with 2D structure. Compared with other deep structures, CNN have got excellent results in image and speech applications. CNN could also be trained by standard back-propagation algorithms. It is easier to train the other depth structures if the parameters have less estimation.

The input image is convolved with three trainable filters. After convolution, three feature maps are generated in the C1 layer. Then, the feature maps are weighted and offset, and then a sigmoid function is used to obtain three S2 level feature maps. These maps are filtered to obtain the C3 layer. This hierarchical structure produces S4 is the same as S2. Eventually, these pixel values are rasterized and concatenated into an input vector to the traditional neural network, and finally the output is obtained. The C layer is the feature extraction layer. The input of each neuron is connected with the local receptive field of the previous layer, and the feature of the local layer is extracted. Once the local feature is extracted, its positional relationship with other features will also fixed. S layer is descending sampling layer. Each computing layer of the network is composed of multiple feature maps, each feature is mapped into a plane, the weights of all neurons in the plane are equal. The sigmoid function with smaller kernel is adopted as the activation function of the convolution network, and the feature mapping is movement invariant.

IV. SYSTEM DESIGN

Cloud-based robots service means more than one robot in a local area network. It is similar with cloud storage and cloud processing mechanisms in computers. All the servers or clients could associate with each other through the so-called cloud. The server or client data are sharing and coordinating a large number of computing tasks. We prepare a main server in a supermarket. All the other robots connect to the main server through the wireless network. The data storage space is allocated and the tasks are scheduled by the main server.

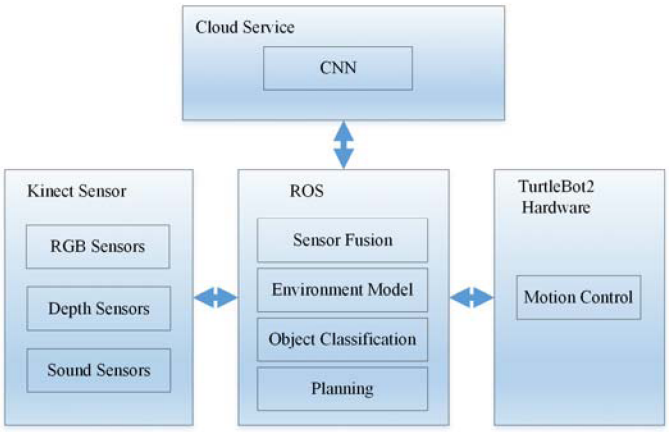

Fig. 3. The Architecture of Shopping Assistant Robot

A. User Cases Design

We have designed five user cases as follows. There are two Kinect sensors on each Robot. Kinect sensor 1 will detect the environment of the Shopping Centre and navigation the robot. Kinect sensor 2 will recognize and track the customer which the robot served for. These two Kinect sensors are installed on two platforms with motor that can rotate 360 degrees. There are 5 user cases (UC) designed by us.

- UC1: Detect of the environment by Kinect Sensor 1. Use the RGB-D images/Video to identify the aisle and the shelf. According to the bone reorganization, recognize the person on the aisle.
- UC2: Detect of the customer by Kinect Sensor 2. There is bone detection and face tracing by Kinect. The Kinect sensor 2 can listen to the simple questions from customer and answer the simple questions.

- UC3: Navigation in the shopping center. Use the Deep learning method to get the location of the Robot and calculate the route to the specific aisle. If there are obstacles before the robot, rotate the Kinect sensor to detect what the obstacle is, after that to determine what next step is (Such as wait for other person pass through first, turn left or right to avoid collision, recalculate the route)
- UC4: trace the customer during the navigation. When Kinect sensor 1 moves to the right places, the Kinect sensor 2 should trace the customer. And the speed will be adjusted to adapt the customer's pace rate. The Kinect sensor 2 could chat with customer during the navigation.
- UC5: Detect the expressions of the customer when the service is finished. Use Kinect sensor 2 to detect the expression of the customer to identify whether they are satisfied by the service.

B. ROS Design

Ubuntu 16.04 is installed on the main control computer. ROS version which we adopted is Kinetic.

We design four main packages in ROS: Sensor Fusion, Environment Model, Object Classification and Planning. Sensor Fusion package is in charge of inquiring and fusion the sensor data. Environment Model package is in charge of processing the environment information, such as creating the map, avoiding obstacle etc. Object Classification package exchanges the data with cloud server. It sends pictures or sound information to cloud server, and gets the result of the classification. Planning package is responsible for the route planning and interacting with TurtleBot2 hardware. The models design is shown as Fig. 4.

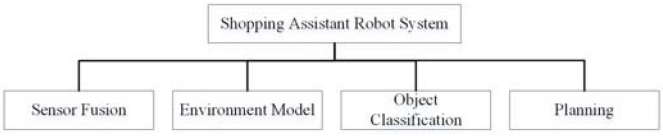

Fig. 4. The Models Design of Shopping Assistant Robot System

V. CONCLUSION

We present a design of shopping assistant robot based on deep learning and ROS. The robot gets the environment information by two Kinect sensors. One sensor is in charge of inquiring the position signal, and takes the pictures of products. The other sensor is in charge of attracting the customer. ROS system is an open source project. There are so many packages for indoor mapping, simultaneous localization and mapping and route planning. The system is designed based on ROS. Finally, Turtlebot2 is used as the hardware platform for the robot.

ACKNOWLEDGMENT

This research is partly supported by the National Natural Science Foundation of China (Grant No. 61501338, 61401308) and Wuhan Chenguang Youth Science and technology program (Grant no. 2014072704011247).

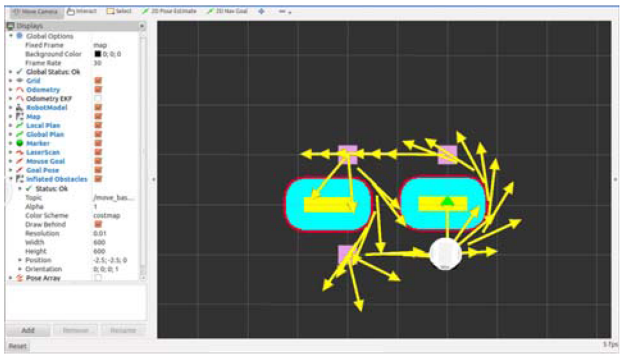

Fig. 5. The Simulation fo Turtlebot in ROS

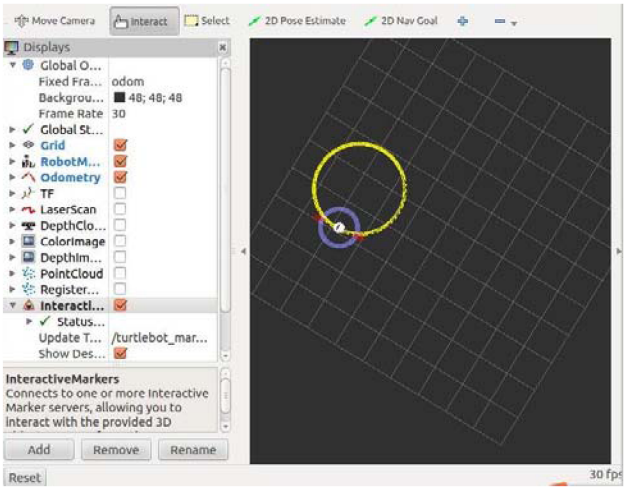

Fig. 6. The Simulation fo Turtlebot in ROS

REFERENCES

[1] D Silver, A Huang, CJ Maddison, A Guez, L Sifre, Mastering the game of Go with deep neural networks and tree search, Nature, vol.529, 2016 pp. 484–489

[2] Y LeCun, Y Bengio, G Hinton, Deep learning, Nature, 2015, pp. 436–444.

[3] Morgan Quigley, Brian Gerkeyy, Ken Conleyy, Josh Fausty, Tully Footey,Jeremy Leibsz, Eric Bergery, Rob Wheelery, Andrew Ng, ROS: an open-source Robot Operating System, ICRA Workshop on Open Source Software, 2009

[4] Rui Min, Neslihan Kose,Jean-Luc Dugelay, KinectFaceDB: A Kinect Database for Face Recognition, IEEE Transactions On Systems, Man, And Cybernetics: Systems, vol. 44, 2014, pp. 1534–1548

[5] G Hinton, S Osindero, YW Teh, A fast learning algorithm for deep belief nets, Neural Computation, vol.18, 2006, pp. 1527–1554

[6] Shuiwang Ji, Wei Xu, Ming Yang, 3D Convolutional Neural Networks for Human Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, 2013, pp. 221–231

[7] N Liu, J Li, Q Liu, H Su, W Wu, Blind source separation using higher order statistics in kernel space, COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol.35, 2016, pp. 289–304