Object Shape and Force Estimation using Deep Learning and Optical Tactile Sensor

Pornthep Sarakon^{a,*}, Yuta Sakai^b, Kazuhiro Shimonomura^b, Hideaki Kawano^{a,*} and Seiichi Serikawa^a

^aDepartment of Electrical and Electronic Engineering, Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan

^bDepartment of Robotics, College of Science and Engineering, Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan

*Corresponding Author: champ.level.m.m@gmail.com, kawano@ecs.kyutech.ac.jp

Abstract

Touch sensing plays a necessarily essential role in robot perception. It helps robot understanding its surrounding environment and in particular the object that it interacts with. For this reasons, robots are equipped with tactile sensor. Moreover, tactile object recognition exhibits a challenge in practical scenarios. In this paper, we proposed object shape classification and force estimation based on deep learning and optical tactile sensor with one touching. This study consists of three methods. First, tactile image is selected by optical flow technique. Image augmentation is used to increase a number of image. Second, features of each image are extracted by modified VGG-16 layer. Last, object shape and force estimation classifier are multi-layer perceptron (MLP), which is a supervised learning technique. The experimental result shows that an accuracy rate is 98.9% for classifying six object shapes and 98.68% for estimating eleven force levels. The results showed that our method outperformed the previous methods that use tactile image and one touching.

Keywords: Tactile sensor, random-dot type optical tactile sensor, VGG-16 layer, MLP.

1. Introduction

In the last decades, a robot plays an important role in industrial. There are several benefits of robots, such as production capacity, safety and work accuracy. These reasons cause robot sensing being much gotten interest by researcher. Visual sensing is a widespread sensing and

many robots are equipped with visual sensors. However, it has problems that not only are light conditions but also occlusions. On the other hand, the other sensing that is also popular and able to solve the problems of visual sensing is touch sensing. Tactile sensor is one of touch sensing. It is a necessarily essential part of robot and is widely used in robot as well.

Currently, there are diverse object recognition studies using tactile sensor. Researches using tactile data based on tactile localization for recognizing object. T. Matsubara, et al.⁽¹⁾ used GPIS models to estimate object shape. G. Vezzani, et al.⁽²⁾ proposed the localization error is minimized by 3D contacted point coordinates.

In addition, there are several researches using spatio-temporal data so as to classify object. H. Yang, et al.⁽³⁾ presented technique based on tactile sequences using LDS with Symmetric Transition matrix. M. Madry, et al.⁽⁴⁾ also used ST-HMP to extract structures from raw tactile data. It did not need to predefine discriminative data characteristics. 3T- Randomized Tiling Convolution Network model was proposed by L. Cao, et al.⁽⁵⁾. It base on feature representation. S.S. Baishya, et al.⁽⁶⁾ studied material classification. By using Fourier spectrum signal, the performance was reached up to 97.3% by CNN.

Besides spatio-temporal data, tactile image data was also used to recognize object by researches. M.M. Zhang, et al.⁽⁷⁾ presented a triangle histogram for 3D tactile recognition. It based on XYZ contact positions on an object surface. A. Schneider, et al.⁽⁸⁾ proposed low-resolution intensity images and k-means clustering for object identification, as well as I. Abraham, et al.⁽⁹⁾ studied a

Fig. 1 Selecting process base on optical flow. Figures (a-d) show optical flow calculating in each frame. Others, (e-d), are raw frames of each optical flow calculating frame.

binary sensor combined with ergodic exploration. There are many researches using data fusion. First, it is visual and tactile data fusion. P. Falco, et al.⁽¹⁰⁾ used visual raw data in training while was testing with tactile raw data. J. Yang, et al.⁽¹¹⁾ also proposed a fusion framework. Tactile sequence was shown by the multivariate times series model and image was characterized by covariance descriptor. Additionally, S. Luo, et al.⁽¹²⁻¹⁵⁾ studied tactile-SIFT using tactile array. By using dictionary generation and histogram representation, the features was well distinguished by kNN. Second, data fusion is visual and haptic data fusion. Y. Gao, et al.⁽¹⁶⁾ used the data fusion for classifying surfaces, which was distinguished by CNN. Haptic exploration for creating tactile image was studied by A.M. Cretu, et al.⁽¹⁷⁾. Third, data fusion is haptic and Electroencephalogram (EEG) fusion. Z. Abderrahmane, et al.⁽¹⁸⁾ proposed a haptic zero shot learning algorithm for recognizing new objects. Besides these fusion data, EEG was combined with tactile signal for classifying object shape⁽¹⁹⁾.

Furthermore, M.V. Liarokapis, et al.⁽²⁰⁾ presented random forests classification technique. It was able to reduce a number of force sensor without significant loss. A. Schmitz, et al.⁽²¹⁾ studied that multimodal object recognition by power grasping of objects with an unknown orientation and position relation to the hand with Deep learning.

Moreover, there are several studies developing a new sensor. C.H. Chuang, et al.⁽²²⁾ presented a high resolution shape recognition based on an ultrasonic tactile sensor. T.

Corradi, et al.^(23,24) developed a new cheap tactile sensor in order to encoding tactile image. The new sensor was capable of distinguishing between object shapes with nearest neighbor. After they improved their technique by using Naive Bayes classifier, accuracy was increased 5%. A kernel sparse coding was developed by H. Liu, et al.⁽²⁵⁻²⁸⁾ in order to address tactile data. By using joint kernel sparse coding (JKSC), they were able to distinguish not only spatial feature but material texture also roughness.

Besides those new tactile sensors, we proposed a random-dot type optical tactile sensor using a camera to acquire tactile information⁽²⁹⁾. By using our tactile sensor's tactile image, we can estimate either shape or force level in one touching and it is different other previous methods that need multi-touching or multi-sensing data to get high performance. In this study, we proposes a new estimation method based on deep learning for classifying either object shape or force level with one touching. The results showed that our method outperformed the previous methods that use tactile image and one touching.

Fig. 2 Object shape. Figures are (a) default, (b) circle, (c) triangle, (d) square, (e) pentagon and (f) hexagon.

2. Object shape classification and Force estimation Method

2.1 Data acquisition

The tactile image acquisition method is processed by the random-dot type optical tactile sensor as shown in Fig. 3. In this sensor, a large number of particle is embedded in a soft gel of a contact part. The tactile image was captured while the contact part was contacting with object that displaced the dots in gel. There are object shape and force level datasets in this study

In object shape classification, raw data is video of circle, triangle, square, pentagon and hexagon shape as shown in Fig. 2 and raw objects are shown as Fig. 3.

The experimental objects

Fig. 3 The tactile sensor. (a) white urethane gel, (b) color sand, (c) transparent urethane gel, (d) transparent acrylic plate, (e) white LEDs, (f) camera.

Fig. 4 Different image data set with different fill-blank modes. (a), (b), (c) and (d) are image in None, Nearest, Reflect and Wrap data set.

Each video is action of pushing object at a random-dot type optical tactile sensor⁽²⁹⁾. Selected frames, shown as Fig. 1, are extracted by overall dense optical flow⁽³⁰⁾ summation being zero as implemented in the OpenCV library. For force estimation, raw data is three images of each eleven force levels as Fig. 5.

However, a number of image in dataset is too little to efficiently classify. Therefore, image augmentation is used to increase a number of image by image processing process. We use Geometric Transformations of Images technique. Although, blank space in rotated image is occurred, it is replaced by fill blank process that is black color, nearest pixel, reflect image and warp image as shown in Fig. 4. By this processing, we can increase a number of image 144 times from beginning.

In sum, our image dataset of object shape are 4 datasets. They are separated by the different fill blank process. Each dataset consists of 6 classes such as default, circle shape, triangle shape, square shape, pentagon shape and hexagon shape. Its number is more enough to be efficiently classify. We separate 66.67, 16.67 and 16.67% (6,912, 1,728 and 1,728 images) for training, validation and test dataset, respectively. Additionally, image dataset force estimation is 1,080 image in each force level. This dataset consists of 11 classes such as 0, 100, 200, 300, 400, 500,

600, 700, 800, 900 and 1,000g. They are separated to the same percentage as dataset of object shape. The final training, validation and test dataset are 7,920, 1,980 and 1,980 images, respectively.

2.2 Feature extraction

Feature extraction is a necessarily initial process. Not only can this process reduce the amount of input data but transform raw data to a set of features, also improve the performance of classification. The reduction of amount of input data can decreases a large amount of memory and computation power requirement. Moreover, Transformation of input data in to a set of features is distinctive properties of input patterns that help differentiating between the

classes of input patterns.

This study chooses Very Deep Convolutional Networks for Large-Scale Visual Recognition proposed by K. Simonyan, et al.⁽³¹⁾, which was created by Visual Geometry Group, University of Oxford. This network, VGG-16 layers, is popular extractor as pre-trained network. The pre-trained network will be connect to classifier (MLP) as shown in Fig. 6. Its main structure consists of convolutional neural networks layers called CNN and fully-connected layers called FC.

Its seventh fully-connected layer, FC7, is used to pull out features from raw image⁽³²⁾. It is the secondly last layer before predicting layer as well as its input and output are 224 x 224 image size and 4,096 features, respectively.

Fig. 6 The experimental network in this study

Although its performance is well, it spends much time and consumes a large amount of memory if we extract features every epoch. We modify a function of FC7, which is removed dropout regularization function, in order to get all of the feature in one time. For this reason, we pull out feature at one time before feed into classifier. Summary, it works well with our classifier.

2.3 Object shape classification and Force estimation

Multi-Layer Perceptron (MLP) is a supervised learning technique. It is well known as neural networks. Its structure seems a group of perceptron that is a single neuron model, and it is created to emulate simple models of biological brain for solving difficult computational tasks.

This study uses MLP as classifier and it is called as model in this study. The model consists of 3 layers, such as input layer, hidden layer and classifier layer, which is presented as Fig. 7. Firstly, the input layer is fixed a number of units at 4,096 units, which equals to a number of features

Fig. 7 The Multi-Layer Perceptron is used in this study

in each image. Secondly, the hidden layer is tested for discovering the best number of its units. Lastly, the classifier layer is the same as the input layer that is fixed, but is added Softmax function. A number of units at classifier layer is 6 or 11 units depending on object shape classification or force estimation.

Moreover, we prevent over fitting in the model with batch normalization and dropout regularization function. The over fitting negatively impacts to the performance of the model on new data. Batch normalization is a technique to provide any layer in a neural network with inputs that are zero mean and unit variance. Dropout regularization is a regularization technique for neural network models proposed by Srivastava N, et al.⁽³³⁾. It is a technique where randomly selected neurons are ignored during training.

In sum, the necessary parameters in the model are batch size and hidden layer units, which is done experiment

in order to get the best training environment. Batch size range is between 32 and 1024, which is range of 2^N . A number of hidden layer units is also range of 2^N . The other parameters are fixed, such as optimizer = Adam, epoch = 2,000, ratio for separating dataset and dropout ratio = 0.5.

3. Experimental result and Discussion

The experiments have been executed using Intel i7 with 3.7 GHz processor, 16 GB RAM, NVIDIA GeForce GTX 1080Ti, Python 3.5 and Chainer library 3.4.0. There are there variables considered in this study, such as batch

size comparison, a number of hidden layer units comparison and fill-blank mode comparison. In object shape classification case, the best parameter is batch size = 1,024, a number of hidden layer units = 256 and fill-blank mode = Warp mode. The best accuracies and losses are come out with Warp data set. Its accuracy is 98.9% and testing time is 6.33×10^{-8} second. In force estimation case, most parameters is as same as the parameters of object shape classification but except a number of hidden layer units = 512. The reason is that a number of classes is different between them and it causes the best hidden layer units of both of them being dissimilar. Accuracy of force estimation is 98.68% and testing time is 1.91×10^{-8} second.

Moreover, the confusion matrix of object shape classification, Table 1, is described in the result of Warp dataset. The confusion matrix presents that the most successful shape is default and failure is circle. If shapes is sorted from greatest to least success, shape list will be default, triangle, square, pentagon, hexagon and circle shape. Almost all failure circle and pentagon images are predicted to hexagon shape. On the contrary, almost every hexagon image is distinguished to pentagon shape. The reason is that raw image of circle, pentagon and hexagon are not been able to distinguish by people because there is much similarity between them.

In addition, the confusion matrix of force estimation is explained in Table 2. The most successful force level is 0g and 1,000g. Nearly all of failure force levels are predicted to nearby force level. For example, failure prediction of 500g class is distinguished to 400g and 600g.

TABLE 1. The confusion matrix of Warp dataset of object shape classification

Truth	Predict					
	default	circle	triangle	square	pentagon	hexagon
default	0.997	0	0	0	0.003	0
circle	0	0.976	0.007	0	0	0.017
triangle	0	0	0.997	0	0	0.003
square	0	0	0	0.993	0.007	0
pentagon	0	0	0	0.003	0.990	0.007
hexagon	0	0.003	0.003	0	0.010	0.983

TABLE 2. The confusion matrix of force estimation

Truth	Predict										
	0g	100g	200g	300g	400g	500g	600g	700g	800g	900g	1000g
0g	1	0	0	0	0	0	0	0	0	0	0
100g	0	0.989	0.011	0	0	0	0	0	0	0	0
200g	0	0.006	0.994	0	0	0	0	0	0	0	0
300g	0	0.006	0	0.989	0.006	0	0	0	0	0	0
400g	0	0	0	0	0.972	0.028	0	0	0	0	0
500g	0	0	0	0	0.011	0.978	0.011	0	0	0	0
600g	0	0	0	0	0	0.006	0.994	0	0	0	0
700g	0	0	0	0	0	0	0.011	0.989	0	0	0
800g	0	0	0	0	0	0	0	0.006	0.978	0.011	0.006
900g	0	0	0	0	0	0	0	0	0.017	0.972	0.011
1000g	0	0	0	0	0	0	0	0	0	0	1

4. Conclusions

This paper proposes object shape classification and force estimation using deep learning and optical tactile sensor. By using tactile image and deep learning, this study is able to efficiently discriminate the different of object shape and force estimation. In object shape case, the accuracy of model is 98.90% and testing time is 6.33×10^{-8} second. The best parameters are batch size = 1,024, a number of hidden units = 256 and fill-blank mode = Warp mode. In force estimation case, the accuracy is 98.68% and testing time is 1.91×10^{-8} second. The best parameters are the same as object shape classification case, but except a number of hidden units = 512. By power of modified VGG-16 layer and our MLP, the result is outperformed the previous methods that use tactile image and one touching. Most failure images are circle, pentagon and hexagon shape because of their much similarity.

However, there is difference between them while moving. This problem will be solved by using video. In addition, if we change color module in raw image for differentiating between background and object, the

performance of model can be improve. Last, the best ratio of dropout regularization finding may increase the efficiency of model.

In future work, we will improve the performance of model with sequence image, color module changing and the best ratio of dropout regularization.

References

- (1) T. Matsubara, K. Shibata, and K. Sugimoto, "Active touch point selection with travel cost in tactile exploration for fast shape estimation of unknown objects", International Conference of Advanced Intelligent Mechatronics (AIM), pp. 1115-1120, 2016.
- (2) G. Vezzani, N. Jamali, U. Pattacini, G. Battistelli, L. Chisci, and L. Natale, "A novel Bayesian filtering approach to tactile object recognition", International Conference of Humanoid Robots (Humanoids), IEEE-RAS, pp. 256-263, 2016.
- (3) H. Yang, F. Sun, W. Huang, L. Cao, and B. Fang, "Tactile sequence based object categorization: A Bag of features modeled by Linear Dynamic System with Symmetric Transition Matrix", International Joint

- Conference of Neural Networks (IJCNN), pp. 5218-5225, 2016.
- (4) M. Madry, L. Bo, D. Kragic, and D. Fox, "ST-HMP: Unsupervised spatio-temporal feature learning for tactile data", International Conference of Robotics and Automation (ICRA), pp. 2262-2269, 2014.
 - (5) L. Cao, R. Kotagiri, F. Sun, H. Li, W. Huang, and Z.M.M. Aye, "Efficient spatio-temporal tactile object recognition with randomized tiling convolutional networks in a hierarchical fusion strategy", Proc. AAAI National Conf. Artif. Intell.(AAAI), pp. 3337-3345, 2016.
 - (6) S.S. Baishya, and B. Bäuml, "Robust material classification with a tactile skin using deep learning", International Conference of Intelligent Robots and Systems (IROS), pp. 8-15, 2016.
 - (7) M.M. Zhang, M.D. Kennedy, M.A. Hsieh, and K. Daniilidis, "A triangle histogram for object classification by tactile sensing", International Conference of Intelligent Robots and Systems (IROS), pp. 4931-4938, 2016.
 - (8) A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features", International Conference of Intelligent Robots and Systems, 2009 (IROS), pp. 243-248, 2009.
 - (9) I. Abraham, A. Prabhakar, M.J. Hartmann, and T.D. Murphey, Ergodic exploration using binary sensing for nonparametric shape estimation, IEEE Robotics and Automation Letters 2, no. 2, pp. 827-834, 2017.
 - (10) P. Falco, S. Lu, A. Cirillo, C. Natale, S. Pirozzi, and D. Lee, "Cross-modal visuo-tactile object recognition using robotic active exploration", International Conference of Robotics and Automation (ICRA), pp. 5273-5280, 2017.
 - (11) J. Yang, H. Liu, F. Sun, and M. Gao, "Object recognition using tactile and image information", International Conference of Robotics and Biomimetics (ROBIO), pp. 1746-1751, 2015.
 - (12) S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel tactile-SIFT descriptor for object shape recognition", IEEE Sensors Journal 15, no. 9, pp. 5001-5009, 2015.
 - (13) S. Luo, X. Liu, K. Althoefer, and H. Liu, "Tactile object recognition with semi-supervised learning", International Conference on Intelligent Robotics and Applications, pp. 15-26. Springer, Cham, 2015.
 - (14) S. Luo, W. Mou, K. Althoefer, and H. Liu, "Iterative Closest Labeled Point for tactile object shape recognition", International Conference of Intelligent Robots and Systems (IROS), pp. 3137-3142, 2016.
 - (15) S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review", Mechatronics 48, pp. 54-67, 2017.
 - (16) Y. Gao, L.A. Hendricks, K.J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data", International Conference of Robotics and Automation, pp. 536-543, 2016.
 - (17) A.M. Cretu, T.E.A. Oliveira, V.P. Fonseca, B. Tawbe, E.M. Petriu, and V.Z. Groza, "Computational intelligence and mechatronics solutions for robotic tactile object recognition", International Symposium of Intelligent Signal Processing (WISP), pp. 1-6, 2015.
 - (18) Z. Abderrahmane, G. Ganesh, A. Crosnier, and A. Cherubini, "Haptic Zero-Shot Learning: Recognition of Objects Never Touched Before", 2017.
 - (19) M. Pal, A. Khasnobish, A. Konar, D.N. Tibarewala, and R. Janarthanan, "Performance enhancement of object shape classification by coupling tactile sensing with EEG", International Conference of Electronics, Communication and Instrumentation (ICECI), pp. 1-4, 2014.
 - (20) M.V. Liarokapis, B. Calli, A.J. Spiers, and A.M. Dollar, "Unplanned, model-free, single grasp object classification with underactuated hands and force sensors", International Conference of Intelligent Robots and Systems (IROS), pp. 5073-5080, 2015.
 - (21) A. Schmitz, Y. Bansho, K. Noda, H. Iwata, T. Ogata, and S. Sugano, "Tactile object recognition using deep learning and dropout", International Conference of Humanoid Robots, pp. 1044-1050, 2014.
 - (22) C.H. Chuang, H.K. Weng, J.W. Chen, and M.O. Shaikh, "Ultrasonic tactile sensor integrated with TFT array for force feedback and shape recognition", Sensors and Actuators A: Physical 271, pp. 348-355, 2018.
 - (23) T. Corradi, P. Hall, and P. Iravani, "Tactile features: recognising touch sensations with a novel and inexpensive tactile sensor", Conference Towards Autonomous Robotic Systems, pp. 163-172. Springer, Cham, 2014.
 - (24) T. Corradi, P. Hall, and P. Iravani, "Bayesian tactile object recognition: learning and recognising objects using a new inexpensive tactile sensor", International Conference of Robotics and Automation, pp. 3909-3914, 2015.
 - (25) H. Liu, D. Guo, and F. Sun, "Object recognition using tactile measurements: Kernel sparse coding methods", IEEE Transactions on Instrumentation and Measurement 65, no. 3, pp. 656-665, 2016.
 - (26) H. Liu, Y. Yu, F. Sun, and J. Gu, "Visual-tactile fusion for object recognition", IEEE Transactions on Automation Science and Engineering 14, no. 2, pp. 996-1008, 2017.
 - (27) H. Liu, J. Qin, F. Sun, and Di Guo, "Extreme kernel sparse learning for tactile object recognition", IEEE transactions on cybernetics 47, no. 12, pp. 4509-4520, 2017.
 - (28) H. Liu, Y. Wu, F. Sun, and D. Guo, "Recent progress on tactile object recognition", International Journal of Advanced Robotic Systems 14, no. 4 : 1729881417717056, 2017.
 - (29) Y. Sakai, P. Sarakon, H. Kawano and K. Shimonomura, "Pattern discrimination using a random dot type optical tactile sensor", Conference of Robotics and Mechanics, 2018.
 - (30) Farnebäck, Gunnar, "Two-frame motion estimation based on polynomial expansion", In Scandinavian conference on Image analysis, Springer, Berlin, Heidelberg, pp. 363-370, 2003.

- (31) K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.
- (32) W. Yuan, C. Zhu, A. Owens, M.A. Srinivasan, and E.H. Adelson, "Shape-independent hardness estimation using deep learning and a GelSight tactile sensor", Robotics and Automation (ICRA), 2017 IEEE International Conference on, pp. 951-958, 2017.
- (33) N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", The Journal of Machine Learning Research 15, no. 1, pp. 1929-1958, 2014.