

Real-time Grasp Type Recognition Using Leap Motion Controller

Yuanyuan Zou^{1, 2}[0000-0002-7309-6643], Honghai Liu³, Jilong Zhang⁴

¹ School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China

² National-Local Joint Engineering Laboratory of NC Machining Equipment and Technology of High-Grade Stone, Shenyang 110168, China
yyzou@sjzu.edu.cn

³ School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK

⁴ Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

Abstract. The recognition of grasp type is essential for a more detailed analysis of human action. In this paper, we propose a novel method for real-time grasp type recognition using Leap motion controller (LMC). Our proposal is based on the tracking data provided by the LMC sensor and a series of feature descriptors are introduced and extracted from LMC data. Combining the feature descriptors of relative positions of thumb, finger joint angles and finger directions lead to the best representation of the arrangement of the fingers. And then the grasp type classification can be achieved by using a SVM classifier. An experimental study of our approach is addressed and we show that recognition rate could be improved. The current implementation is also can satisfy the real-time requirements.

Keywords: Grasp Type Recognition, Feature Representation, SVM, Leap Motion Controller.

1 Introduction

Human hand action recognition is the key for many research areas such as human computer interaction, robot control, robotic action planning, etc., and is attracting more and more research interest. Understanding human hand usage is one of the richest information source to recognize manipulation actions [1]. Human uses various types of hand grasp in daily living activities and industrial manipulations. Hand grasp types are important for understanding hand manipulation since they characterize how hands hold the objects during manipulation [2]. Therefore, the recognition of grasp types is essential for a more detailed analysis of human action. Various methods of the grasp recognition have been explored, including EMG-based [3] and vision-based [4] etc. Among these methods, vision-based methods are used widely since they can be less cumbersome uncomfortable than the other methods due to no physical contact with users.

Leap motion controller (LMC) is a recently introduced sensor which is based on vision techniques for hand gesture recognition [5]-[7]. Comparing to depth cameras

2

such as the Microsoft Kinect [8], LMC is working on a smaller 3D region and extracts data with higher accuracy [9]. It can be exploited for accurate gesture recognition. In this paper, we present a real-time grasp types recognition approach using LMC, where the main advantage is the extraction of a set of relevant features that can effectively recognize the grasp types.

This paper is structured as follows. Section II presents the architecture of the proposed hand grasp type recognition system and the grasp types selected for recognition in this work. Our recognition approach is described in Section III. The following section highlights the experimental results. The paper is enclosed by a conclusion and future perspectives.

2 System Description

2.1 System Architecture

The general architecture of the system presented in this paper is shown in Fig. 1. The LMC is used for capturing the hand gestures. The LMC is composed of two cameras and three LEDs, which can infer the 3D position of a subject's hand. The features are extracted from 3D hand data and then are processed with the classification method of Support Vector Machines (SVM).

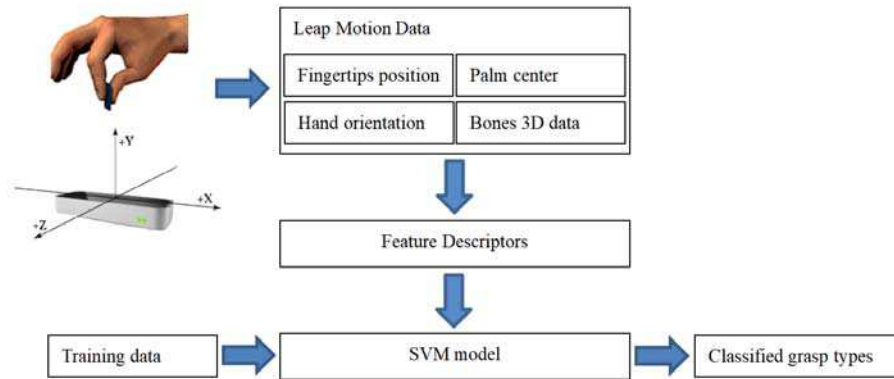


Fig. 1. Proposed hand grasp type recognition system architecture.

2.2 Grasp Types

The grasp taxonomy of human hand grasp type is important for understanding the way human grasp objects and knowing common use patterns. A number of work have investigated the categorization of hand grasp. Feix et al. [10] compares the existing taxonomies and extracts 33 grasp types which are arranged in a novel taxonomy according to the number of fingers in contact with the object and the position of the thumb. Heumer et al. [11] identifies six different grasp types, including power, palm-up, hook, oblique, precision and pinch which are sufficient for grasping strategies in daily living activities.

In this work, we select six types of hand grasps based on functionality which are usually in daily life activities and industrial in-hand manipulations. The list of selected grasp types is shown in Fig. 2.

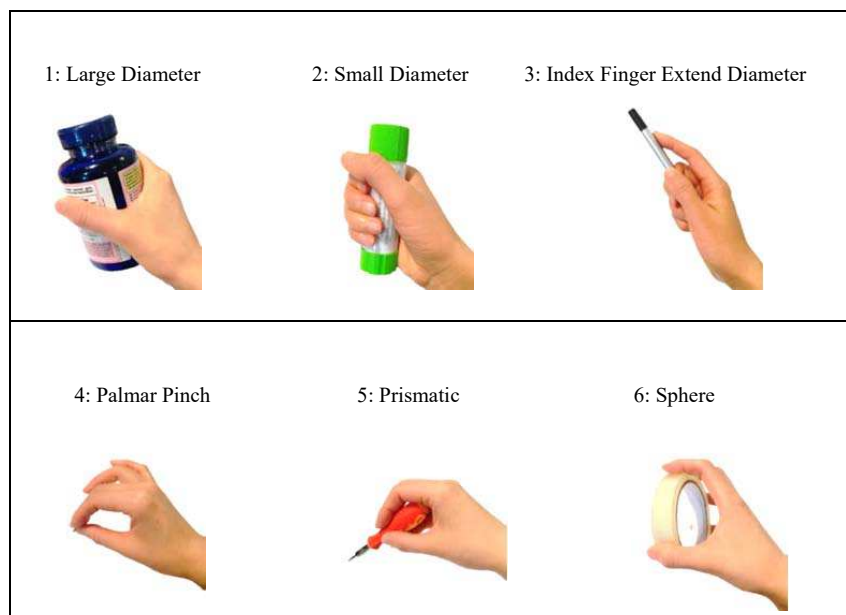


Fig. 2. The list of selected grasp types.

3 Proposed Approach

3.1 Feature Representation

Feature extraction is the major step in vision based hand gesture recognition system. To represent the characteristics of hand gestures, various features have been extracted. Many approaches have been presented and mostly are based on the standard scheme of extracting relevant features from the RGB depth data [12]. Hand shape are represented with Histogram of Oriented Gradient (HOG) features which are image descriptors based on collected local distributions of intensity gradients. Cai et al. extracts three HOG features for palm regions which encodes the shape of different hand postures [13]. The LMC captures the skeletal data of hand gesture and mainly provide the following data such as Number of detected fingers, Position of the fingertips, Palm center, Hand orientation, Hand radius, Bones data etc. Some data acquired by LMC are shown in Fig. 3. The device provides directly some of the most relevant points for gesture recognition and allows avoiding complex computations needed for their extraction from depth and color data.

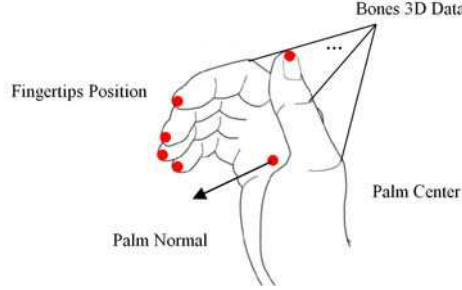


Fig. 3. Data acquired by LMC.

However, these absolute positions of the hand and fingers are not relevant as gesture features, but they can be used to obtain other meaningful features. Therefore, to represent the hand grasp types in Fig. 2 entirely, we extract 5 features, including the feature descriptors of Fingertip distances proposed by [14], the feature descriptors of Extended index finger representation, Relative positions of thumb and Finger joint angles proposed in this paper, and Finger directions.

Let i be the index of a finger, Vectors \mathbf{P}_i containing the 3D positions of each of the detected fingertips, the center of the hand palm is \mathbf{C} , the normal to the plane that corresponds to the palm region pointing downward from the palm center is \mathbf{n} .

Fingertip distances: The fingertip distances are defined as the 3D distance of the fingertips from the palm center. The vector of finger i is computed by:

$$\mathbf{v}_i = \mathbf{P}_i - \mathbf{C}, \quad i = 1, \dots, 5 \quad (1)$$

$$d_i = \|\mathbf{P}_i - \mathbf{C}\| / S_1, \quad i = 1, \dots, 5 \quad (2)$$

Where S_1 is the scale factor as shown in equation (3), which is to make the approach robust to people with hands of different size. It is defined as the maximum distance between the palm center and the fingertips. The feature values are normalized in the interval $[0, 1]$ by dividing the value.

$$S_1 = \max(\|\mathbf{P}_i - \mathbf{C}\|) \quad (3)$$

Then the first feature vector is built as follows:

$$\mathbf{f}_1 = [d_1, d_2, d_3, d_4, d_5] \quad (4)$$

Extended index finger binary representation: The Extended index finger binary representation is defined as if the index finger is extended for the grasp gesture.

$$eb = \begin{cases} 1, & \text{extended} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The second feature vector is built as follows:

$$\mathbf{f}_2 = eb \quad (6)$$

Relative positions of thumb: The relative positions of thumb are defined as the 3D distances of the fingertips from the thumb fingertip.

$$dr_i = \|\mathbf{P}_1 - \mathbf{P}_i\| / S_2, \quad i = 2, \dots, 5 \quad (7)$$

5

Where S_2 is also a scale factor as shown in equation (8) which is defined as the maximum value of the distances of the other fingertips from the thumb fingertip. The feature values are normalized in the interval $[0, 1]$ by dividing the value.

$$S_2 = \max(\|\mathbf{P}_1 - \mathbf{P}_i\|) \quad (8)$$

Then the third feature vector is built as follows:

$$f_3 = [dr_2, dr_3, dr_4, dr_5] \quad (9)$$

Finger joint angles: The Finger joint angles are defined as the angles of every finger joint. The LMC can capture bones 3D data, so the joint angles of each finger can be calculated respectively. For three joints are present on each finger, with the exception of the thumb, which possesses only two. Therefore, the fourth feature vector is built as follows:

$$f_4 = \begin{bmatrix} 0, c_{11}, c_{12} \\ c_{21}, c_{22}, c_{23} \\ c_{31}, c_{32}, c_{33} \\ c_{41}, c_{42}, c_{43} \\ c_{51}, c_{52}, c_{53} \end{bmatrix} \quad (10)$$

Finger directions: The finger directions are identified as the angles between the finger vector and the normal.

$$a_i = \angle(\mathbf{v}_i, \mathbf{n}), i = 1, \dots, 5 \quad (11)$$

Then the fifth feature vector is built as follows:

$$f_5 = [a_1, a_2, a_3, a_4, a_5] \quad (12)$$

3.2 Grasp Types Classification

The Support Vector Machine (SVM) is used as the classifier for grasp types classification. The SVM is a powerful tool for classification and is widely used in object detection and recognition [15].

The SVM model is established by using the different feature descriptors as the input of the network. In order to realize the accurate recognition under the condition of small sample as far as possible, cross validation test is used. The training samples are divided into n parts, where m copies as the part retained, the remaining $n-m$ as parts of training. The training part is used to probability estimation as training set, and the retained part is used to test as test set. The parameters to be chosen mainly include penalty factor and kernel function parameter, which are determined by the method of grid search.

4 Experiments and Analysis

The experiments are designed and conducted to evaluate the proposed approach. We acquire a set of grasp gesture type (GGT) database with a LMC device. It consists of 6 different gestures performed by 6 different subjects a total of 20 times each for 720 different samples. LMC images of grasp gestures in GGT database from different people are shown in Fig. 4.

6

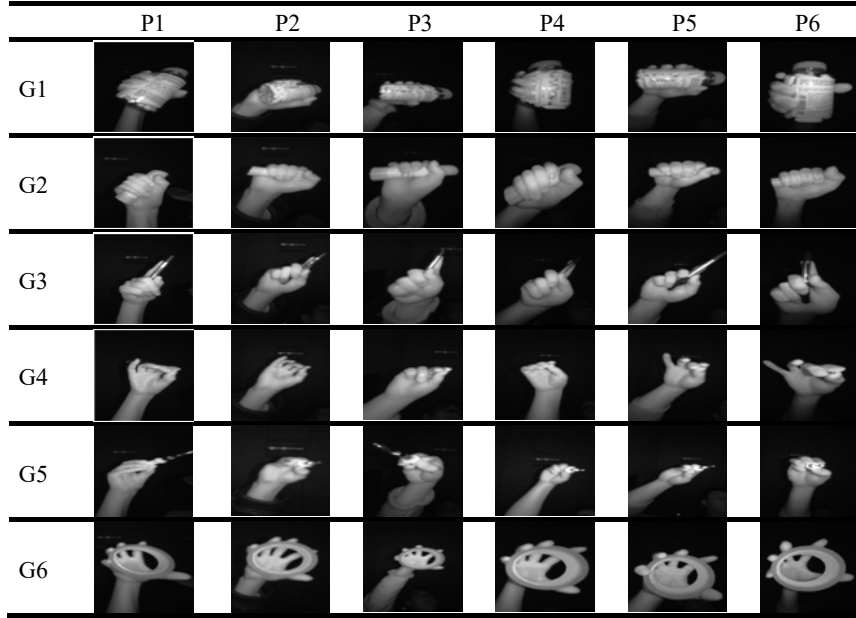


Fig. 4. LMC images of grasp gestures in GGT database.

We split the dataset in a train set and a test set. We place in the train set the data from all the subjects except one and in the test set the data from the remaining subject. To assess the effectiveness of our approach to classify the grasp gesture of GGT dataset into 6 classes, we execute 6 completely independent tests changing the subject each time in the test set. In each test, a train set with 5 persons and a test set with a single person are used. The final recognition accuracy is obtained by averaging the results of the 6 tests.

For all following experiments, we compare the results by using each feature descriptor independently and by reconstructing different feature descriptors. The Table 1 presents the results obtained by using different descriptors.

Table 1. Results of using different descriptors.

Features	Accuracy
f_1	66.72
f_2	30.60
f_3	53.12
f_4	73.15
f_5	65.09
$f_1+f_3+f_4$	74.95
$f_1+f_4+f_5$	72.87

7

$f_3+f_4+f_5$	75.65
$f_1+f_3+f_4+f_5$	74.12
$f_1+f_2+f_3+f_4+f_5$	73.56

From Table 1, we observe that the finger joint angles give the better representation of the arrangement of the fingers and allow obtaining an accuracy of 73.15%. Fingertip distances and finger directions have slightly lower performance. The fingertip distance allow obtaining an accuracy of 66.72% and the finger directions allow obtaining an accuracy of 65.09%. As shown in Table 1, combining the feature descriptors of relative positions of thumb, finger joint angles and finger directions lead to the best representation of the arrangement of the fingers and allow obtaining an accuracy of 75.65%. To better understand the behavior of our approach according to the recognition per class, the confusion matrix from $f_3+f_4+f_5$ is shown in Table 2. The results show that gesture Large Diameter is difficult to be classified correctly and it has a wide confusion with the gesture Small Diameter. The reason is that the fingers maybe shielded by the object when grab a large diameter object, so the sensor is not able to recognize the fingers.

Table 2. The confusion matrix of the proposed approach for GGT database

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆
G ₁	4	15	1	0	0	0
G ₂	3	17	0	0	0	0
G ₃	1	2	17	0	0	0
G ₄	0	0	0	20	0	0
G ₅	1	0	0	2	17	0
G ₆	0	0	0	0	0	20

The experiments are implemented in C# on a PC with an Intel I5-3210M processor and 8GB of RAM. The extraction of features is fast and requires 10us per frame. Finally SVM classification is performed in just 7us per frame. Therefore, the proposed approach is particularly suitable for real time gesture recognition.

5 Conclusions

In this paper, we mainly presented a real-time grasp type recognition system using LMC and investigated the feature extraction of LMC tracking data for selected six types of hand grasp. We proposed the new feature descriptors called relative positions of thumb and finger joint angles which can describe the hand grasp type more effectively. Finally, the experimental results have shown a higher recognition accuracy can be obtained by combining the two new features with finger directions.

As future work, the approach can be studied to be able to utilize two hands and to recognize dynamic grasp. In addition, skeleton-based features can be combined with the sensor images to provide more effective description and to obtain better recognition.

Acknowledgements. The authors would like to acknowledge the support from the Natural Science Foundation of China under Grant No. 51405481 and the Natural Science Foundation of Liaoning Province under Grant No. 20180551124.

References

1. Kim, J. W., You, S., Ji, S. H. and Kim, H. S.: Real-time hand grasp recognition using weakly supervised two-stage convolutional neural networks for understanding manipulation actions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 10-11 (2017).
2. Yang, Y., Guha, A., Fermuller, C. and Aloimonos, Y.: A cognitive system for understanding human manipulation actions. *Advances in Cognitive Systems*, 3, 67-86 (2014).
3. Kakoty, N. M. and Hazarika, S. M.: Recognition of grasp types through principal components of DWT based EMG features. In: 2011 IEEE International Conference on Rehabilitation Robotics, 1-6 (2011).
4. Cai, M. J., Kitani, K. M., Sato, Y.: A scalable approach for understanding the visual structures of hand grasps. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), 1360-1366 (2015).
5. Khelil, B., Amiri, H.: Hand gesture recognition using leap motion controller for recognition of arabic sign language. In: Proceedings of the 3rd International Conference on Automation, Control, Engineering and Computer Science (ACECS'16), 20-22 (2016).
6. De Smedt, Q., Wannous, H., Vandeborre, J.: Skeleton-based dynamic hand gesture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1-9 (2016).
7. Bachmann, D., Weichert, F., Rinkenauer, G.: Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors*, 18(7), 2194-2232 (2018).
8. Du, H. and To, T.: Hand gesture recognition using Kinect, Boston University, (2011).
9. Guna, J., Jakus, G., Pogačnik, M., et al.: An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, 14(2), 3702-3720 (2014).
10. Feix, T., Romero, J., Schmiedmayer, H., et al.: The grasp taxonomy of human grasp types. *IEEE Transactions on Human-machine Systems*, 46(1), 66-77 (2016).
11. Heumer, G., Amor, H. B., Jung, B.: Grasp recognition for uncalibrated data gloves: a machine learning approach. *Presence: Teleoperators and Virtual Environments*, 17(2), 121-142, (2008).
12. Yang, Y., Fermuller, C., Li, Y., et al.: Grasp type revisited: a modern perspective on a classical feature for vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 400-408, (2015).
13. Cai, M. J., Kitani, K. M. and Sato, Y.: Understanding hand-object manipulation with grasp types and object attributes. In: *Robotics: Science and Systems*, (2016).

14. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with jointly calibrated leap motion and depth sensor. *Multimedia Tools and Applications*, 75(22), 14991-15015 (2016).
15. Dutt, V., Chaudhry, V., Khan, I.: Pattern recognition: an overview. *American Journal of Intelligent Systems*, 2(1), 23-27 (2012).