Rotation and Translation Invariant Object Recognition with Tactile Sensors While Grasping

Shan Luo*, *Student Member, IEEE*, Kaspar Althoefer, *Member, IEEE*, Hongbin Liu *Member, IEEE*

***Abstract*—This paper presents a novel approach to recognize objects with feature descriptors invariant to movement and rotation of objects in hands during grasping. As an object is manipulated by hands without prior knowledge, tactile feedback can make up for the information loss caused by vision occlusion. But frequent movement and manipulations make it difficult to recognize shape and pose of the object. To solve this problem, Scale-Invariant Feature Transform (SIFT) technique from the computer vision is used to extract descriptors to represent rotation and translation invariant local features. The proposed system is validated by classifying objects with real data from tactile sensors mounted to a Barrett hand; average overall accuracy of 91.2% has been achieved.**

I. Introduction

Humans employ multiple sensing streams to grasp an object, i.e. vision, tactile sensation, proprioception etc. Among these sensing modalities, rich haptic information from fingers and palm assist us to detect, recognize and manipulate objects in hands with ease, even when vision is occluded by hands. For robots, we also envision such capacity. Earlier researchers have contributed their efforts: Y. Bekiroglu et al. used tactile sensing to make grasp planning to assess grasp stability [1]; J. Ilonen et al created an optimal estimation approach to fuse visual and tactile information with an iterative extended Kalman filter for grasping [2]. However, it is still challenging to determine the identity and pose of objects due to frequent manipulations in during grasping like movement or rotation.

To solve this problem in robotic grasping, we propose to use feature descriptors in tactile sensing. Feature detection and matching have been widely probed in many applications of computer vision, especially for object recognition and tracking. For recognition, keypoint features or so-called interest points are often used to represent objects and recognize different objects. For object tracking, feature correspondences across different images are utilized to trace one or multiple objects. This method can also be applied to tactile modality. With the increasing spatial resolution and spatiotemporal response in last few decades, tactile sensors show the potential to serve as an “imaging” device: for example, in [3] a tactile image based recognition method is proposed. As tactile sensor interacts with objects, tactile imprints can reflect their shapes via pressure distribution. Based on these tactile images, feature descriptors can be obtained, which should be rotation and translation invariant. Scale Invariant Feature Transform (SIFT) descriptors [4] are thus selected because of its extreme good performance in visual object identification invariant to rotation, scales and translation. A bag-of-features framework is then employed and a vocabulary of *k* words is therefore learned by *k*-means unsupervised learning and the histogram codebook is used to identify the unknown object. A previous similar type of approach is described in [5]; however, for each tactile reading, its columns were concatenated to form a descriptor directly. Due to this essence, the recognition is not rotation and translation invariant, hence, an identical object observed at different poses is recognized into different classes, which is not acceptable in grasping scenario. To validate the proposed system, real objects are classified with readings from tactile sensors and a high overall accuracy has been achieved. Fig. 1 depicts our experimental system: a Barrett hand with tactile sensors is employed to perform the grasping manipulations; some object samples are also presented.

Shan Luo, Kaspar Althoefer, Hongbin Liu are with the Centre for Robotics Research, Department of Informatics, King’s College London, London, WC2R 2LS, UK (e-mail: shan.luo@kcl.ac.uk; k.althoefer@kcl.ac.uk; hongbin.liu@kcl.ac.uk).

* indicate the corresponding author { email: shan.luo@kcl.ac.uk }.

Figure 1. Depiction of the system to grasp objects. Left: object pool. Right: Barrett hand with tactile sensors.

II. Methodology and Experimental evaluation

To test the SIFT descriptors in tactile scenario, we take an object of simple shape for testing, i.e. an Allen key. And for simplicity, only the tactile imprint from one finger is considered. The tactile reading of it and subsequent processing are shown in Fig. 2. Interest points (here is a corner) are first detected which represent key features of the object and provide useful information for grasping. Based on their orientation and local appearance, feature descriptors will be formed. For each SIFT descriptor, we get a 128-element

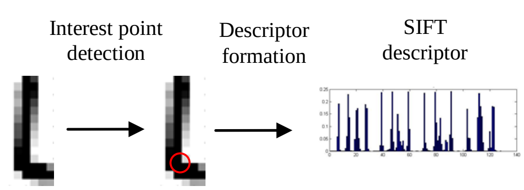

Figure 2. Formation of SIFT descriptors.

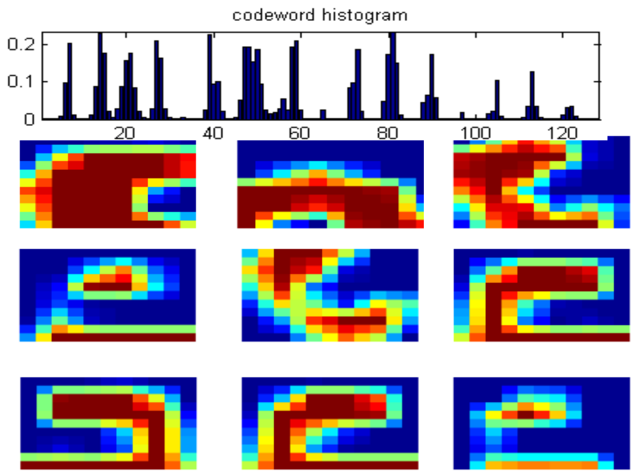

Figure 3. Sub-patches assigned to the same codeword.

histogram.

To perform a recognition task using these descriptors, a bag-of-words framework is adapted, treating the features of objects as words. This framework is employed because it is simple but powerful in classifying objects. A *k*-means clustering algorithm is first applied on the training dataset to create a dictionary of codewords (cluster centroids). Some tactile patches whose descriptors are assigned to the same codeword in the experimental evaluation are illustrated in Fig. 3. It can be noticed that a semicircle appears in all the sub-patches but at different positions, orientations and even are of different intensities. It shows that a codeword is clustered by the descriptors of sub-patches regardless of how these similar features appear in them. In this way, the object recognition can be achieved invariant to movement and rotation of objects during grasping in hands. Histograms of word occurrences for each object are then generated and robot can use these distributions to recognize a large set of various objects by grasping the objects a few times at different positions and comparing the occurrence histogram of the unknown object with the histograms in the database with a *k*-Nearest Neighbors algorithm. A diagram of this framework is illustrated in Fig. 4.

The system is also evaluated by real data. An average overall classification accuracy of 91.2% is achieved with a dictionary size of 50 codewords and 10 times of grasping for each object. It proves the robustness of our algorithm with regards to the randomness of poses and relative positions between objects and robotic hands during grasping.

III. CONCLUSION AND DISCUSSION

In this paper a novel approach is proposed to recognize various objects invariant to their movement and rotation with tactile sensing during grasping. The proposed system is validated by classifying real objects with readings from grasping with tactile sensors attached to a Barrett hand; high classification accuracy has been achieved.

There are still several issues to be solved. First, the resolution of the tactile sensors is still too low, which means that it will bring some problems if we take techniques directly from vision side. Second, even though SIFT are widely used in vision and it performs extremely well, but problems still exist. How to create a good feature descriptor that fit to tactile images is essential. At last, in real conditions there are many interest points for one object. Thus how to select most useful feature is still an issue worth to be explored.

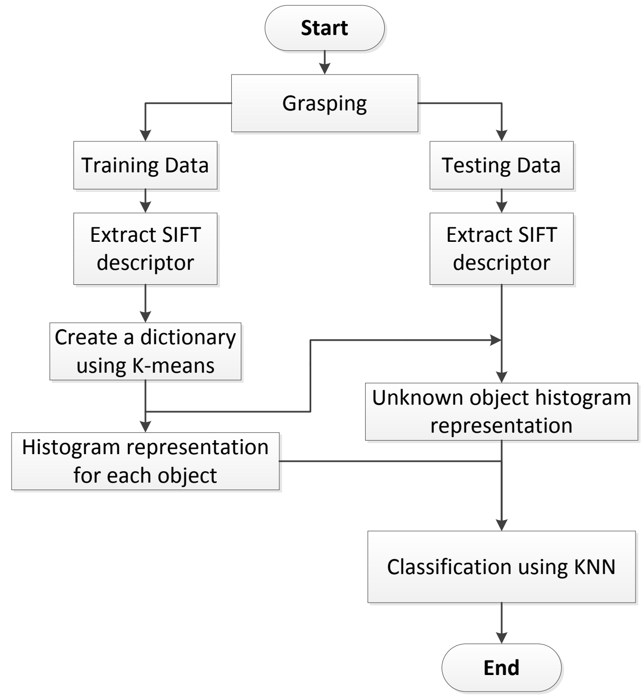

Figure 4. The bag-of-words framework and recognition process to classify unknown objects.

ACKNOWLEDGMENT

The work described in this paper was partially funded by the European Commission's Seventh Framework Programme under grant agreement 287728 in the framework of EU project STIFF-FLOP, and the China Scholarship Council.

REFERENCES

[1] Bekiroglu, Yasemin, Kai Huebner, and Danica Kragic. "Integrating grasp planning with online stability assessment using tactile sensing." In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 4750-4755. IEEE, 2011.

[2] Ilonen, Jarmo, Jeannette Bohg, and Ville Kyrki. "Fusing visual and tactile sensing for 3-D object reconstruction while grasping." In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 3547-3554. IEEE, 2013.

[3] Y.-H. Liu, Y.-T. Hsiao, W.-T. Cheng, Y.-C. Liu, and J.-Y. Su, "Low-Resolution Tactile Image Recognition for Automated Robotic Assembly Using Kernel PCA-Based Feature Fusion and Multiple Kernel Learning-Based Support Vector Machine," *Math. Probl. Eng.*, vol. 2014, pp. 1–11, 2014.

[4] Lowe, David G. "Distinctive image features from scale-invariant keypoints."*International journal of computer vision* 60, no. 2 (2004): 91-110.

[5] Schneider, Alexander, Jürgen Sturm, Cyrill Stachniss, Marco Reisert, Hans Burkhardt, and Wolfram Burgard. "Object identification with tactile sensors using bag-of-features." In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pp. 243-248. IEEE, 2009.