Rotation and Translation Invariant Object Recognition with a Tactile Sensor

Shan Luo*, Wenxuan Mou†, Min Li*, Kaspar Althoefer*, Hongbin Liu*
Email: shan.luo@kcl.ac.uk w.mou@se13.qmul.ac.uk {min.m.li, k.althoefer, hongbin.liu}@kcl.ac.uk
*Centre for Robotics Research, Department of Informatics, King's College London, WC2R 2LS, UK
†School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

***Abstract*—In this paper a novel approach is proposed to recognise different objects invariant to their translation and rotation by utilising a tactile sensor attached to a robotic arm. As the sensor is small compared to the tested objects, the robot needs to access those objects multiple times at different positions and is prone to move or rotate them, that inevitably increases difficulty in object recognition during manipulations. To solve this problem, it is proposed to extract tactile translation and rotation invariant local features to represent objects; a dictionary of *k* words is therefore learned by *k*-means unsupervised learning and a histogram codebook is then used to identify objects. The proposed system has been validated by classifying real objects with data from an off-the-shelf tactile sensor. The average overall accuracy of 91.2% has been achieved with only 10 touches and a dictionary size of 50 clusters.**

I. Introduction

Human beings can recognise objects with ease through our cutaneous sensation. Inspired from this, robots are also envisioned to perform this task via tactile sensors. However, low force and spatial resolution of the tactile sensor hinders its development. To ease the effect of sparse tactile readings, early researchers focused on creating clouds of contact points with tactile sensors to reconstruct the profiles of touched objects. Therefore techniques of computer graphics were widely employed [1] [2]. Allen et al. [1] fitted resultant points from readings of tactile sensors to super-quadric surfaces to build object models and a similar process was conducted in [2] where tensor B-spline surfaces were used instead. Later, researchers investigated approaches to recover local geometry, i.e., surface normals and curvatures, paying much attention to extracting information from each contact point. Fearing et al. [3] used a nonlinear model-based inversion to determine the curvature with a cylindrical tactile sensor. And in [4] it is proposed to describe a patch through polynomial fitting under an estimated Darboux frame determined by two principal directions and surface normals at the curve intersection points. Some other researchers distinguished contact shapes by employing machine learning, which can be summarised in two steps: 1. Features are first extracted from local contact shapes. 2. A classifier is then trained to classify shapes. In [5], by covariance analysis of pressure values in tactile images, three orthogonal axes were acquired and a Naïve Bayes classifier was used to recognise local object features. There is also some work been done to recover the global image of observed objects using tactile sensors, e.g., Pezzementi et al. [6] proposed a mosaic method to synthesise local geometric surfaces to recover the object-level geometric surface using histogram and particle filters, in which the objects were a set of raised letters. Recent studies also consider other object perceptual properties. Xu et al. [7] employed multimodal tactile sensors to recognise objects with a Bayesian exploration procedure whereas Madry et al. [8] probed in utilising temporal tactile measurements to recognise objects.

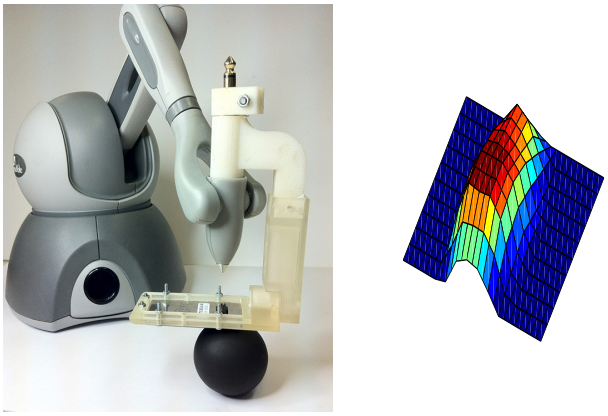


Fig. 1: Depiction of the system to recognise objects, i.e., a soft ball here, with an array tactile sensor attached to a Phantom Omni manipulator arm. Left: experimental set-up. Right: tactile readings.

In this paper, it is proposed to recognise objects invariant to their translation and rotation with tactile sensing in a framework of Bag-of-Words (BoW). This framework was first used in tactile scenarios by Schneider et al. [9].However, tactile readings were taken as features directly, that makes identical objects observed at different poses being recognised as different identities. Pezzementi et al. [10] took one step further: multiple descriptors were extracted from tactile images and compared to each other. However, a considerable number of samples (around 50) were needed to achieve reasonable classification results. Compared to the above work, the novelty of our method is as follows: 1. Rotation and translation invariant descriptors are extracted from tactile images, that enables a robot to recognise objects when it moves or rotates them. 2. The required number of contacts is reduced but high

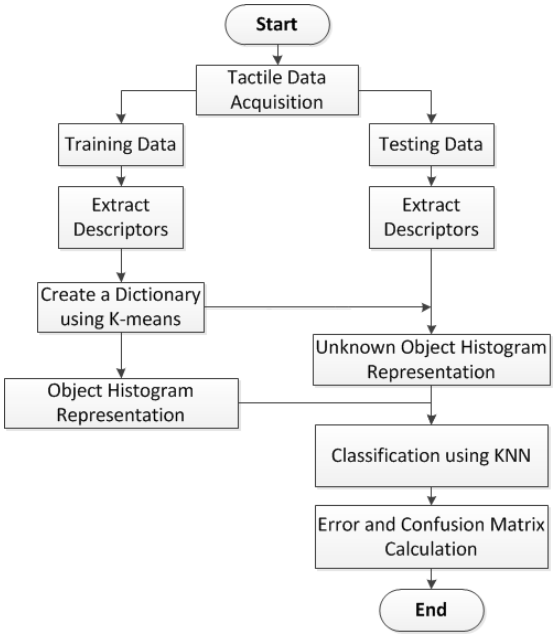


Fig. 2: The bag-of-words framework and recognition process to classify unknown objects.

accuracy can still be achieved. The proposed system has been validated by identifying real objects with readings from an off-the-shelf tactile sensor and a high overall accuracy was achieved. Figure 1 shows the experimental system and sampled patches from the tactile sensor.

II. METHODOLOGY

A. Overview

As the robot finger is smaller than the tested objects $\boldsymbol{O}=\{o_1,o_2,\ldots,o_n\}$, only limited surface area of an object can be touched by the tactile sensor. Therefore only partial local information can be perceived. In our system, the local observations of each object are the acquired tactile patches $\boldsymbol{W}=\{w_1,w_2,\ldots,w_n\}$, which present in normalised pressure values of the sensing elements organised in a matrix form. To perform a global classification based on these local image patches, a bag-of-words framework from computer vision [11] is adapted to treat the features of objects as words. It is a simple but powerful technique to classify objects. Given the low-resolution intensity images recorded with the tactile sensor, the descriptors of these images are extracted and a dictionary is then generated from the training dataset by $\boldsymbol{k}$-means clustering. Histograms of word occurrences for object classes are then created and robot can use these distributions to identify an object by touching it a few times at different positions and comparing its occurrence histogram with the histograms in the database (Figure 2).

B. Feature quantisation

Inspired by Scale Invariant Feature Transform (SIFT) descriptors [12] from computer vision, it is proposed to use image gradient directions to form descriptors $\boldsymbol{P}=\{p_1,p_2,\ldots,p_n\}$.

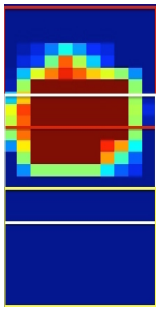


Fig. 3: A regular grid of sub-patches.

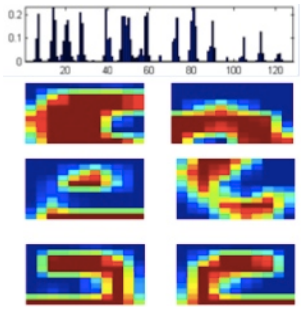


Fig. 4: Sub-patches assigned to a codeword.

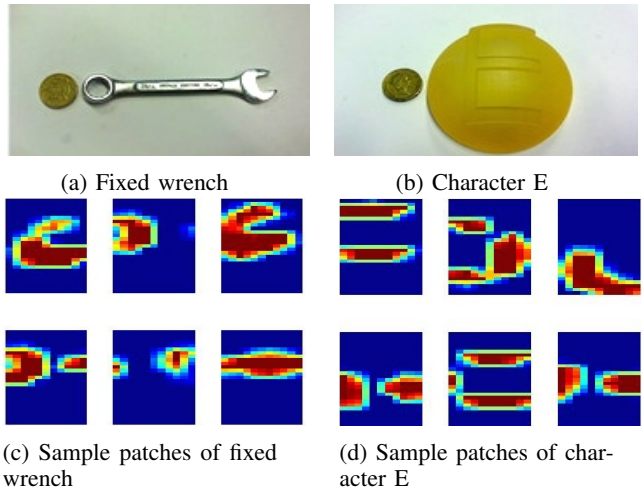
Unlike using camera vision for shape recognisation, tactile sensing allows mapping real dimensions of pressed objects. Therefore tactile images do not need to be scaled. In visual images, key points such as corners are viewed as distinctive features. And multiple key points can be detected in one image due to affluent information. However, in each tactile image there is limited information present and such features are much less. Therefore key point localisation is eliminated as in [13]. To make features more robust, we divide each tactile image into three equivalent regions and extract one 128 dimensional SIFT descriptor for each region as shown in Figure 3, taking region centres as "key points".

C. Dictionary generation and object representation

These descriptors are then clustered to "codewords", which are similar to words in text documents. Therefore this produces a "codebook", which is similar to a dictionary. This means that a codeword can be considered as a representative of several similar descriptors. As the dictionary may vary with different object databases, unsupervised $\boldsymbol{k}$-means clustering is employed and the learned cluster centroids $\boldsymbol{c}$ are obtained as codewords, in which Euclidean distances between descriptors $\boldsymbol{p}$ and codewords $\boldsymbol{c}$ are calculated as in (1). Some sub-patches whose descriptors are assigned to the same codeword in the experimental evaluation are illustrated in Figure 4. It can be noticed that a semicircle appears in each but at different positions and orientations. It shows that a codeword is clustered regardless of how these features appear. In this way, the object recognition can be achieved invariant to movement and rotation of objects. The objects are then represented as occurrence histograms $\boldsymbol{h^o}$ with $\boldsymbol{k}$ bins in total. Each bin is with an initialised value 0 and added one when a descriptor is assigned to it. $\boldsymbol{h^o}$ is normalised at last by L norm as shown in (2).

$$d(p_i, c_i) = \sum_{k=1}^{128} |p_i(k) - c_i(k)| \tag{1}$$

$$h_i^o \leftarrow \frac{h_i^o}{\sum_{i=1}^{n} |h_i^o|} \tag{2}$$



(a) Fixed wrench (b) Character E

(c) Sample patches of fixed wrench (d) Sample patches of character E

Fig. 5: Objects with sample patches.

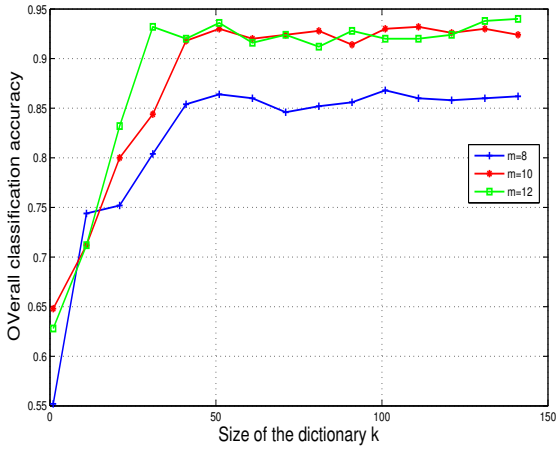


Fig. 6: Overall accuracies with various dictionary sizes.

D. Classification using *k*NN

The ***k***-Nearest Neighbour (***k***NN) classifier is employed to classify objects in our system. Here the number of neighbours ***k*** is set to 1. The similarity between histograms of test objects and objects in the database is computed using histogram intersection as in (3).

$$d(h^{test}, h^{class}) = 1 - \sum_{i=1}^{k_{dict}} min(h_i^{test}, h_i^{class}) \qquad (3)$$

III. EXPERIMENTAL EVALUATION

A. Experimental Setup

A resistive Weiss tactile array sensor is attached to the stylus of a haptic device Phantom Omni, which serves as a robotic manipulator. The tactile sensor consists of 84 sensor cells located in 14 rows and 6 columns with a size of 51 mm $\times$ 24 mm as a whole and a spatial resolution of 3.4 mm for each cell. The sensor is covered by elastic rubber foam to conduct the externally applied force,which is sampled at a rate of 5 frames per second. The raw readings were preprocessed in two steps: 1). If in a tactile image the maximum value is lower than the specific threshold or the sum of all elements is smaller than a predefined decision value, it is considered as collected unintentionally and deleted. 2). The readings were then normalised, hence, falling into [0, 1].

The data acquisition is carried out as follows: 1. An idle load is initialised with no interaction; this serves as a reference measurement. 2. For every object, the exploring procedure is repeated 5 times and during each the stylus is controlled to move 1 cm and rotate in 4 orientations by step, keeping the planar surface of the sensor normal to the object surface; in this way, the entire object surface is covered. The first four times were taken as the training set while the last one is taken as the test set. To verify that only a few touches are needed to recognise objects, ***m*** patches in the last procedure were sampled randomly for each test set. As a result, 2500 tactile images for 10 objects were collected. The objects were all taken from the lab environment or daily life (fixed wrench, wooden cuboid, plier, wheel model, wrench, hook, coffee cup, soft ball and comb) with an exception of a 3D printed character E on a hemisphere, which possesses three dimensional features. Figure 5 shows a fixed wrench and a character E and their corresponding sampled patches (the tactile images were interpolated for visualisation but in the processing raw data were used).

B. BoW model building

To create a dictionary of local features, the descriptors were first extracted. The SIFT descriptors were calculated with a sample sub-patch size of six and a grid spacing of three. Thereby three sub-patches and corresponding three 128-element descriptors were generated for one 14×6 tactile image. Each 6×6 sub-patch was adapted into a grid of 4×4 bins and gradient distributions in 8 orientations (from 0 to 1.75π by 0.25π) in each bin were summed thus a 128 dimensional descriptor was obtained for each sub-patch. These descriptors were then clustered by the ***k*** means algorithm to form a dictionary.

C. Evaluation

It is apparent that the larger the size of the dictionary ***k*** is, the higher the accuracy of the recognition of objects. As the touches in the test set were randomly selected, the accuracy of the recognition was repeated ten times for each dictionary size and mean values of these ten trials were calculated. The effect of the increase of ***k*** can be seen from Figure 6 with 8, 10 and 12 patches utilised respectively. It is evident that the accuracy increases as ***k*** grows but it levels off when the size is greater than 50. The likely reason for this is that "synonyms" will happen if the size increases more.Thus a dictionary size of 50 was chosen.

The effect of the number of touches ***m*** on the recognition performance was also investigated. As for the dictionary size, ten trails were taken and the mean values were calculated

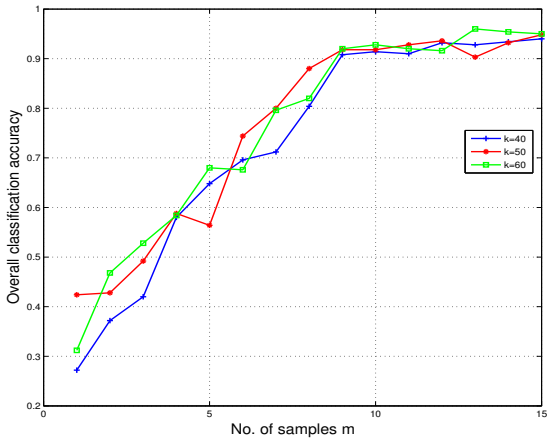


Fig. 7: Overall accuracise with different number of samples.

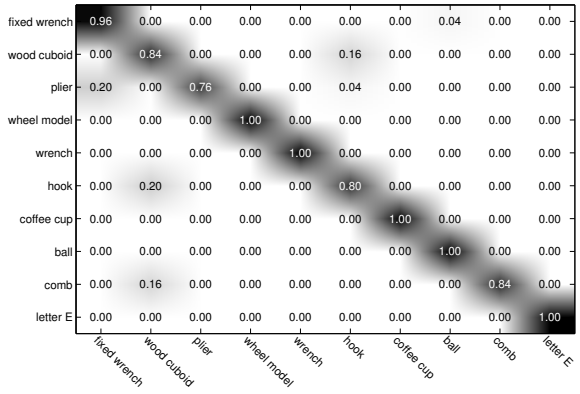


Fig. 8: Confusion matrix of object recognition.

to get a better certainty. The dictionary sizes of 40, 50 and 60 were being used for comparison. Figure 7 shows that the more times the robot touched the objects, the more probable it became for the objects to be classified correctly. But reasonable accuracy could be obtained when ten samples were collected hence the robot only needs a few observations to reach a reasonable guess.

D. Classification results

Based on the discussion, the dictionary size $\boldsymbol{k}$=50 and touch times $\boldsymbol{m}$=10. An average overall classification accuracy of 91.2% was achieved. The confusion matrix of the experiments shown in Figure 8. It proves the robustness of our algorithm with regards to different poses and relative positions between objects and the tactile sensor. On the other hand, some of the objects were assigned to wrong labels, i.e., some observations of the cuboid were wrongly concluded to be from the hook and vice versa. This was caused by their common features such as linear lines.

IV. Conclusions and Future Work

In this paper it is proposed to recognise objects invariant to their movement and rotation with a tactile sensor by using rotation and translation invariant local features. A vocabulary of $\boldsymbol{k}$ words is learned by $\boldsymbol{k}$-means unsupervised learning and the histogram codebook is used to identify objects by $\boldsymbol{k}$NN. The proposed system was validated with an off-the-shelf tactile sensor and high classification accuracy was achieved. This work has many potential applications such as robotic grasping. In future work, the positions of the tactile sensor will be considered to involve more spatial geometric information, potentially allowing to classify more complicated objects. And the Support Vector Machine (SVM) classifier could be an alternative for the $\boldsymbol{k}$-NN classifier.

Acknowledgment

This work was supported 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] P. K. Allen and P. Michelman, "Acquisition and interpretation of 3-d sensor data from touch," in *Interpretation of 3D Scenes, 1989. Proceedings., Workshop on*. IEEE, 1989, pp. 33–40.

[2] M. Charlebois, K. Gupta, and S. Payandeh, "Shape description of general, curved surfaces using tactile sensing and surface normal information," in *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, vol. 4. IEEE, 1997, pp. 2819–2824.

[3] R. S. Fearing and T. O. Binford, "Using a cylindrical tactile sensor for determining curvature," *Robotics and Automation, IEEE Transactions on*, vol. 7, no. 6, pp. 806–817, 1991.

[4] Y.-B. Jia, L. Mi, and J. Tian, "Surface patch reconstruction via curve sampling," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 1371–1377.

[5] H. Liu, X. Song, T. Nanayakkara, L. D. Seneviratne, and K. Althoefer, "A computationally fast algorithm for local contact shape and pose classification using a tactile array sensor," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1410–1415.

[6] Z. Pezzementi, C. Reyda, and G. D. Hager, "Object mapping, recognition, and localization from tactile geometry," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5942–5948.

[7] D. Xu, G. E. Loeb, and J. A. Fishel, "Tactile identification of objects using bayesian exploration," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3056–3061.

[8] M. Madry, L. Bo, D. Kragic, and D. Fox, "St-hmp: Unsupervised spatio-temporal feature learning for tactile data," in *IEEE International Conference on Robotics and Automation (ICRA)(to appear)*, 2014.

[9] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 243–248.

[10] Z. Pezzementi, E. Plaku, C. Reyda, and G. D. Hager, "Tactile-object recognition from appearance information," *Robotics, IEEE Transactions on*, vol. 27, no. 3, pp. 473–487, 2011.

[11] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 604–610.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification via plsa," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 517–530.