Novel Tactile-SIFT Descriptor for Object Shape Recognition

Shan Luo, Wenxuan Mou, Kaspar Althoefer, Hongbin Liu*

***Abstract*—Using a tactile array sensor to recognize an object often requires multiple touches at different positions. This process is prone to move or rotate the object, which inevitably increases difficulty in object recognition. To cope with the unknown object movement, this paper proposes a new Tactile-SIFT descriptor to extract features in view of gradients in the tactile image to represent objects, to allow the features being invariant to object translation and rotation. The Tactile-SIFT segments a tactile image into overlapping sub-patches, each of which is represented using a *dn*-dimensional gradient vector, similar to the classic SIFT descriptor. Tactile-SIFT descriptors obtained from multiple touches form a dictionary of *k* words, and the Bag-of-Words method is then used to identify objects. The proposed method has been validated by classifying 18 real objects with data from an off-the-shelf tactile sensor. The parameters of the Tactile-SIFT descriptor, including the dimension size *dn* and the number of sub-patches *sp*, are studied. It is found that the optimal performance is obtained using an 8-dimensional descriptor with 3 sub-patches, taking both the classification accuracy and time efficiency into consideration. By employing Tactile-SIFT, a recognition rate of 91.33% has been achieved with a dictionary size of 50 clusters using only 15 touches.**

***Index Terms*— Tactile sensors, object recognition, robot tactile systems.**

I. Introduction

The sense of touch is irreplaceable for us human beings, especially when we explore the environment in close vicinity when vision is occluded. It conveys the sensory information, i.e., pressure, vibration, pain, temperature etc., to our central nervous system, therefore assisting us to perceive the ambient world and avoid potential injuries. Research has exhibited that, compared to vision and audition, the tactile sensations demonstrate superiority at processing the material characteristics and detailed shapes of objects [1]. As humans, robots are also expected to possess the tactioception. To achieve this goal, there is a rapid expansion of tactile sensor development using different sensing principles in last few decades [2]–[7]. In contrast, however, the research in decoding the conveyed tactile information is still in the early stage. The material characteristics and shapes of objects are the two main objectives to be revealed by the tactile sensation. Some researchers have focused on identifying material properties [8]–[10]. Decherchi *et al.* take multiple techniques from computational intelligence to classify object materials with tactile data [9]. Liu *et al.* [10] apply a dynamic friction model to determine the surface physical properties while a robotic finger slides along the object surface with a varying sliding velocity.

To recognize contact shapes, one approach is to recover local geometry from each contact point, i.e., surface normal and curvature. By using a cylindrical tactile sensor, Fearing *et al.* propose a nonlinear model-based inversion to recover contact surface curvatures [11]. Jia *et al.* analyze one-dimensional tactile data to describe a patch through polynomial fitting under an estimated Darboux frame determined by two principal directions and surface normal at the curve intersection point [12]. Clouds of contact points have also been used to reconstruct object shapes thanks to techniques of computer graphics [13]–[16]. Allen *et al.* fit resultant points from tactile sensors readings to super-quadric surfaces to reconstruct unknown shapes [15] and a similar process is conducted in [17] but tensor B-spline surfaces are used instead. Through these methods, arbitrary contact shapes can be identified by estimating surface curvatures. However, the investigation of large object surface using this method can be time consuming and key features are not revealed.

Another approach to recognize contact shapes is to use sensors with tactile arrays. One method is to employ machine learning techniques and it can be divided into two steps: 1) features are extracted from the gained tactile readings of objects with known contact shapes; 2) a classifier is then trained and applied to predict the shapes of test objects. By covariance analysis of pressure values in tactile readings, it is proposed in [18] to acquire three orthogonal principal axes, namely, eigenvectors of the covariance matrix of the pressure pattern; a Naïve Bayes classifier is fed with resultant axe lengths, main axe direction and shape convexity to recognize local object features. They also contribute a similar recognition process by transforming each tactile reading into a 512-feature vector and using a neural network classifier in [19]. The neural networks are also used in [20] to classify local shapes but it is found to be sensitive to pattern variations in positions and orientations.

An alternative approach is to treat tactile arrays as images and apply vision descriptors to extract object features. Inspired by the similarities between tactile patterns and grey-scale images, Ji *et al.* [21] transform tactile data into histograms as structured features to discriminate four basic human-robot touch patterns, i.e., poking, a full finger contact, gripping with three fingers and grasping with whole hand. Research has also

Manuscripst received December 16, 2014. This work was partially supported by the China Scholarship Council. Part of this work has been presented in IEEE Sensor Conference, Valencia, Spain, 2014.

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; hongbin.liu@kcl.ac.uk; k.althoefer@kcl.ac.uk).

Wenxuan Mou is with the Computer Vision Research Group, the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK (e-mail: w.mou@qmul.ac.uk).

* indicates the corresponding author {email: hongbin.liu@kcl.ac.uk }.

been done to explore suitable features for the tactile recognition. In [22], the columns of each tactile matrix are concatenated to form a vector that is directly treated as a descriptor. Due to its essence, the recognition is not invariant to object movements. Therefore, one identical object is recognized as distinct identities if it is placed in different orientations to the robotic gripper. The regional descriptors are utilized in [23] to recognize object shapes, but objects can only be classified into four classes, i.e., planar, one-edged, two-edged and cylindrical objects. The kernel PCA-based feature fusion is used in [24] to transform geometric features and Fourier descriptors into ones to better discriminate objects but only geometric shapes can be classified in this work. In [25], several descriptors from computer vision are employed and compared to achieve a satisfactory classification performance but a considerable quantity of contacts (around 50) is required. The Scale Invariant Feature Transform (SIFT) descriptor is created based on image gradients [26]. It has been proved that in computer vision applications the SIFT is robust to cope with the object's pose variation. This is particularly useful for object recognition using tactile images, since the touch could introduce unexpected object rotation and translation. SIFT has been initially explored in [25] for tactile object recognition, however, it did not show good performance. The main reason could be due to that the design of the original SIFT is overcomplicated for processing tactile images. The high dimensional descriptor of the original SIFT tends to over-fit the lower dimensional tactile images. To overcome this problem, we propose a novel Tactile-SIFT descriptor which is suitable for processing the tactile images by reformatting the SIFT as follows: 1) scale pyramid building and key-point localization are eliminated; 2) the descriptor dimension is reduced to find an appropriate descriptor to achieve a high classification accuracy and good time efficiency at the same time; 3) each tactile image is segmented into overlapping sub-patches and a descriptor is extracted from each sub-patch.

To recover the global identity of the object using the obtained local tactile images, several approaches have been proposed. Pezzementi *et al.* utilize tactile images to characterize local geometric surfaces and propose a mosaic method to synthesize these patches to recover the object-level surface using both histogram and particle filters [27], in which the objects are a set of raised letters. The tactile sensors are utilized to collect data formed as contact point clouds in [28] and a probabilistic model is built to classify objects. However, both these two approaches require large number of contacts. Another approach is to use the unsupervised learning to generate a codebook of tactile features for the object [21], [22], [25]. Among these methods, the Bag-of-Words (BoW) model, originally widely applied in image categorisation [29], has been proved to be easy for implementation and have high recognition accuracy. Thus, in this paper the BoW model is selected for the global tactile object recognition.

The original design of Tactile-SIFT descriptor proposed in our previous work [30] has been further modified and studied in this paper. The contributions of this paper can be summarised as follows. 1) Tactile-SIFT descriptors with various dimensions are created and compared, and it is found that a high accuracy can be reached by using 8-dimensional Tactile-SIFT descriptors. Thus the 8-dim descriptors are chosen to be used in the tactile object shape recognition. 2) The methods to segment tactile images are also investigated and it is discovered that a good performance can be achieved when a tactile image is divided into three sub-patches with overlaps between each other. 3) 18 real objects are involved in the experiments. To the best knowledge of the authors, this is the largest real object sample size for studying tactile object recognition to date. The results demonstrate that the proposed Tactile-SIFT descriptor is suitable for tactile object recognition and a high recognition rate can be reached with a few touches. To conduct this study, a test rig was developed as shown in Figure 1. This test rig consists of a Weiss Robotics tactile array sensor WTS 0614-34 with 6×14 sensing elements and a Sensable PHANTOM Omni device for tactile sensor positioning.

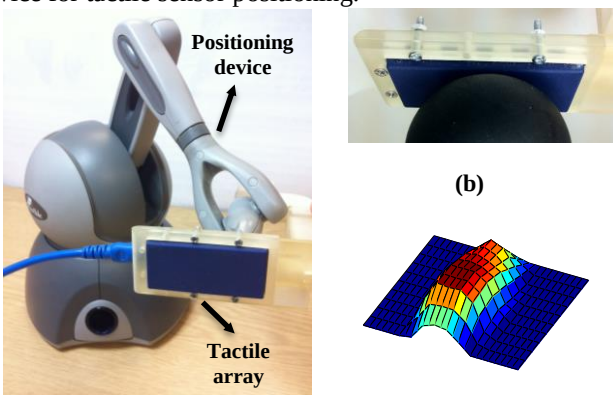


Fig. 1. (a) Depiction of the system to recognize objects, with a tactile array sensor attached to a manipulator arm. (b) Tactile sensor and object interaction. (c) Tactile readings.

II. Methodology

A. Overview

In general cases, the robotic fingertips are smaller than the objects $O = \{o_1, o_2, \cdots, o_n\}$ that they get in contact with. Thus only limited surface area of an object can be accessed by the "skin" of the fingertips during the manipulations in hand, which means that the objects are needed to be touched multiple times to obtain a global image. In our proposed system, the local observations of each object are acquired as tactile patches of the same size, which are presented in normalized pressure values of the sensing elements organized in a matrix form. To perform a global classification based on them, a BoW framework is employed and adapted, treating the tactual features of objects as words. A data flow is illustrated in Figure 2 and it works in the following way: 1) given the collected low-resolution tactile images, the Tactile-SIFT descriptors (features) $[p_1, p_2, \cdots, p_n]$ are extracted; 2) a dictionary $W = [w_1, w_2, \cdots, w_k]$ is then generated from these descriptors by k-means clustering; 3) histograms of word occurrences for object classes are then generated based on the features of the training dataset. Thus the robot can use these distributions to identify an object while touching it a few times at different positions and comparing its occurrence histogram with the histograms in the database.

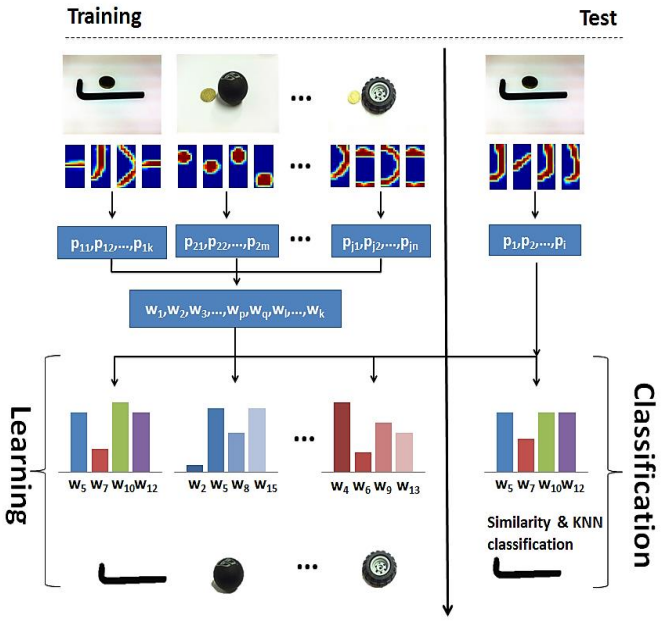


Fig. 2. The bag-of-words framework and recognition process to classify unknown objects.

B. *Feature quantization*

We use words in our dictionary to describe explored objects and the words represent partial features of an object to form its monolithic image. For instance, with descriptions of "round", "sleek", "hard" and "thin handle" we can probably conclude an object as coffee mug. In the same way, in the robotic tactile perception, descriptors can also be extracted to represent essential information in a form of numerical vectors. To make descriptors invariant to object movements, we propose to adapt Scale-Invariant Feature Transform (SIFT) [26] descriptors to tactile object recognition, which have performed very well in the object recognition using visual images.

In classic SIFT algorithm, a scale pyramid is first built to make descriptors scale invariant. However, unlike using camera vision for shape recognition, tactile sensing allows mapping real dimensions of pressed objects. Therefore, tactile images do not need to be scaled; that means that the scale pyramid needs not be built. Besides, in visual images key points such as corners are viewed as distinctive features and multiple can be detected in one image due to affluent information. But in each tactile image there is limited information present and such features are much less. Due to this reason, key point localisation is also eliminated and patch centre is taken as "key point" instead, as the way used in scene classification [31].

To make features more robust, it is proposed to segment each tactile image into several overlapping sub-patches and one descriptor is extracted for each. The sizes of sub-patches are all the same for different partitions. Two examples are illustrated in Figure 3. In Figure 3(a), the tactile image is segmented into three sub-patches and thus three descriptors p_1, p_2, p_3 are obtained; in Figure 3(b), keeping the sub-patch size same as the case of Figure 3(a), the tactile image is interpolated along the vertical axis first and the resulted image is segmented into four sub-patches and four descriptors p_1, p_2, p_3, p_4 are obtained.

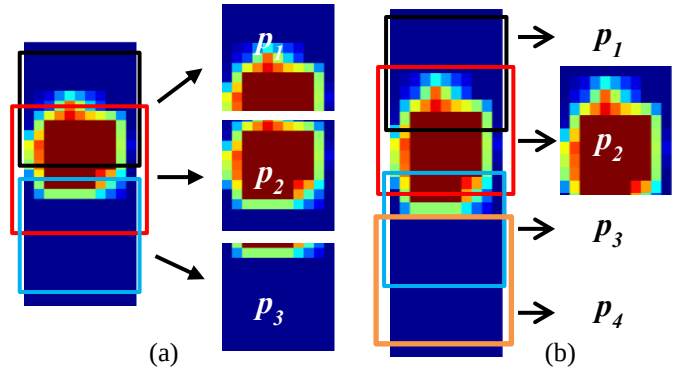


Fig. 3. Examples of a regular grid of (a) three and (b) four sub-patches for a tactile reading. And one descriptor is obtained from each sub-patch.

The descriptor for each sub-patch is computed as follows. First, the gradient magnitude and orientation are calculated by using a difference-of-Gaussian function at each pixel of the 6×6 sub-patch. To mitigate abrupt changes in the image, a Gaussian weighting function is applied independently to blur each sub-patch [26]. Second, the sub-patch is divided into a grid of cells and three sizes of grids, i.e., 4×4, 2×2, 1×1, are employed. Third, an orientation histogram is created for each grid cell from the gradient orientations of pixels within or intersect with the cell, which has 8 or 4 bins covering the 360° range of orientations. The value of each bin is computed by summing products of each gradient magnitude of the pixels within or intersect with the grid cell and a weight (= 1-d), where d is the distance of the pixel to the center of the grid cell measured in units of the grid cell spacing. At last, all the orientation histogram entries are concentrated into a vector and the descriptor is obtained from normalizing the vector by the L2 vector norm. As a result, descriptors of different dimensions, i.e., 4×4×8=128, 4×4×4=64, 2×2×8=32, 2×2×4=16, 1×1×8=8 and 1×1×4=4, are created.

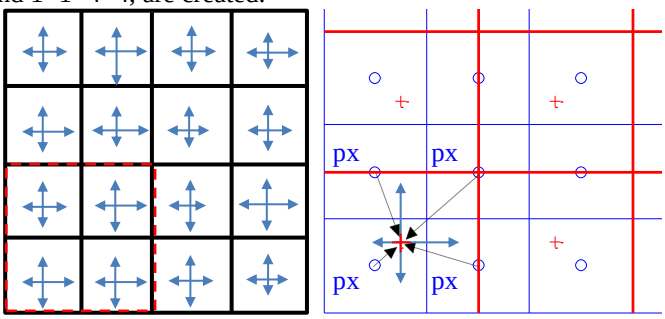


Fig. 4. The 64-dim descriptor obtained from a 6×6 sub-patch tactile image. (a) The sub-patch is divided into 4×4 grid cells, in each of which an orientation histogram of 4 bins is obtained. All these histogram entries form the 64-dim descriptor. (b) The enlarged bottom left region of the sub-patch, marked by the dashed box in (a); grid of cells are marked with red lines, pixels and grid cell centers are represented by blue circles and red crosses respectively. The orientation histogram formation of the grid cell is also illustrated, which is contributed by the gradients of four pixels within or intersect with this grid cell, labelled as *px*. Black arrows denote the distance of each pixel to the cell center.

The formation of a 64-dim descriptor is illustrated in Figure 4. A 6×6 sub-patch is divided into 4×4 grid cells, shown in Figure 4(a). The value of each bin, i.e., the length along that direction, is computed as follows: 1) each gradient vector of these four pixels is projected onto that direction; 2) the

projections are multiplied by the associated weights; 3) the products are summed up to obtain the length of that direction. Figure 4(b) illustrates the computational procedures. An orientation histogram of 4 bins is obtained for each grid cell of the sub-patch and all the histogram entries, the lengths of the arrows, are allocated into an array to form a 64-dim descriptor.

C. Dictionary generation and histogram representation

After the descriptors are created, they are clustered to obtain "codewords" that are similar to words in text documents. Therefore, this produces a "codebook", similar to a dictionary. It means that a codeword can be considered as a representative of several similar descriptors. A set of clusters is obtained through unsupervised *k*-means clustering of training data, *k* refers to dictionary size, and the learned cluster centroids *c* are gained as codewords. Euclidean distances are used to compute the distances between descriptors *p* and codewords *c* as in (1).

$$d(p,c) = \|p - c\| \tag{1}$$

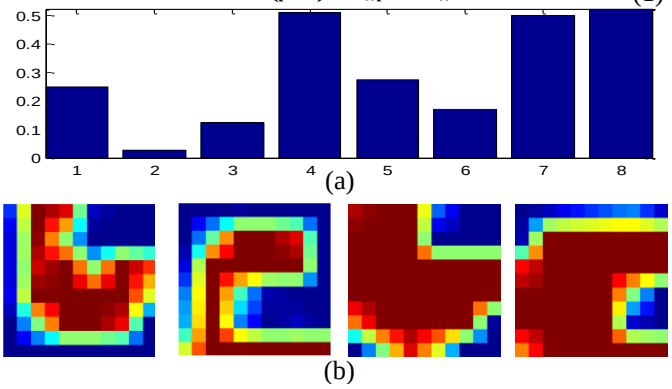


Fig. 5. Sample sub-patches assigned to an 8-dim codeword. (a) The vector representation of the codeword. (b) Sample sub-patches assigned to it.

Some sub-patches are illustrated in Figure 5, whose descriptors are assigned to the same codeword. It can be noticed that a semicircle appears in each but at different positions and orientations. It shows that an 8-dim codeword is clustered regardless of how these features appear. In this way, the object recognition can be achieved invariant to object movements. The objects are then represented as occurrence histograms h^o of codewords with *k* bins in total. Each bin is initialized with a value 0 and is incremented by one when a descriptor is assigned to it. h^o is at last normalized by L1 norm as shown in (2).

$$h^o := \frac{h^o}{\sum_{i=1}^{k} h_i^o} \tag{2}$$

D. Classification using kNN

The *k*-Nearest Neighbor (*k*NN) classifier is employed to classify objects in our system. Here the number of neighbors *k* is set to 1. For every object, the exploration procedure is repeated five times and the data of the first four exploration trials and the last one are taken as training set and test set respectively. To make the learned histograms more reliable, the training data for each object are taken together to form one occurrence histogram based on the codebook. The similarity between histograms of test objects and objects in the database are computed using histogram intersection [32] as in (3). Some other state-of-the-art classifiers such as Support Vector Machine (SVM) have also been studied in the experiments, but as only one histogram vector is obtained to represent each object in the training process, the classification results appear to be over-fitting. Thus *k*NN is employed at the end and the framework works well in the experiments.

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

E. Exploration strategies

It is expected that a better exploration strategy improves the recognition rates [22] but to test the robustness of our algorithm to the variance of the relative positions and poses between the robot and observed objects, a uniformed exploration strategy is employed in our experimental study instead of an informed strategy and it has been shown that a good recognition rate can still be achieved using our algorithm.

III. Experimental Setup

In the test rig, a resistive Weiss tactile sensor is attached to the stylus of Phantom Omni from Sensable that serves as a robotic manipulator. The tactile sensor consists of 6×14 sensor cells with a size of 51 mm by 24 mm as a whole and 3.4 mm by 3.4 mm for each cell. The sensor is covered by elastic rubber foam to conduct the externally applied force. Though the maximum scanning rate is 270 frame/s, the sensor signals are sampled at a rate of 5 frame/s that is found to be sufficient based on our initial studies. The raw readings are pre-processed as follows: 1) if in a tactile image the maximum value or the sum of all the readings is lower than the predefined thresholds, it is considered to be collected unintentionally and deleted; 2) taking the non-linear sensor characteristics into consideration, the readings are normalized by the maximum value of each tactile image to achieve consistency in the dynamic range of collected tactile measurements, falling into [0, 1].

The data acquisition is carried out as follows: 1) an idle load is served as a reference measurement with no object-sensor interaction; 2) for every object, the exploration procedure is repeated five times and during each the stylus is moved manually with a speed of around 5 mm/s, keeping the sensor plane normal to the object surface; in this way, the entire surface of the object is covered while a number of tactile patches are collected. To verify the robustness w.r.t variance of poses, the relative object-sensor poses are chosen randomly. As a result, 7290 tactile images in total for 18 objects, shown in Figure 6, are collected. And the number of collected tactile images for every object is also listed, which varies with the object dimension. The data of first four exploration procedures the last one are taken as training set and test set respectively. A leave-one-out cross validation is carried out by utilizing different dataset as test data to validate the results and a similar performance has been observed. To verify that only a few touches are needed to recognize objects, *m* readings in the test set are sampled randomly to form a test trial. And to get a reliable conclusion, 25 test trials are carried out for each object in one classification procedure in which a same dictionary is used. The classification results for one object are averaged as the

Fig. 6. Objects used for tests. First, third and fifth rows are visual images of objects. The name and number of collected tactile readings are also listed under the picture of each object. Second, fourth and last rows are corresponding sampled tactile images, in which prominent features can be observed.

individual accuracy and overall performance can be computed.

The objects in the experiments, illustrated in Figure 6, are taken from either lab environment or daily life (fixed wrench, wooden cuboid, plier, wheel model, wrench, Allen key, coffee cup, soft ball, comb, mouse, tape, saws framework, tweezers, plug, scissors and wide fixed wrench) with two exceptions, i.e., 3D printed point array on a flatbed and a character E on a hemisphere. Figure 6 also shows the corresponding sampled patches for each object (the tactile images are interpolated for visualization but in the processing raw data are used). As seen in the tactile images, relative sensor-object poses and positions are randomly selected as the same features appear at different orientations and positions in multiple patches of one object.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

There are four variables considered to be studied, i.e., descriptor dimension *dn*, sub-patch number for one tactile reading *sp*, dictionary size *k* and the number of touches *m*. As a whole, it is aimed to achieve high classification accuracy while the number of samples needed for recognition is minimized and time efficiency is increased.

1) Dimensions of descriptors dn

The descriptor dimension results from both the number of grid cells and the number of gradient directions. In this case, *dn* is investigated and varied whereas *sp*, *k* and *m* are fixed. Three descriptors are obtained from each tactile reading (*sp*=3), namely, each image is segmented into three sub-patches; the dictionary size *k* is set to 50; for each test trial *m*=15 samples are taken. The classification procedure (25 test trials each) is carried out 10 times and the average recognition rates are shown in Figure 7. It can be noticed that the descriptors by counting gradient orientations to 8 directions (dimensions 128, 32 and 8) outperform ones with 4 directions (dimensions 64, 16 and 4). It means that the increased division of the directions contributes to increment of the classification accuracy. On the other side, among those with 8 directions, 8-dim descriptors perform best, though 128-dim descriptors outperform 32-dim ones. In the cross validation tests, same results are also demonstrated. The probable reason is that the increase of the grid division brings noises to the classification and the count of 8 directions in the whole tactile image is enough to extract the local information embedded in the low resolution tactile image.

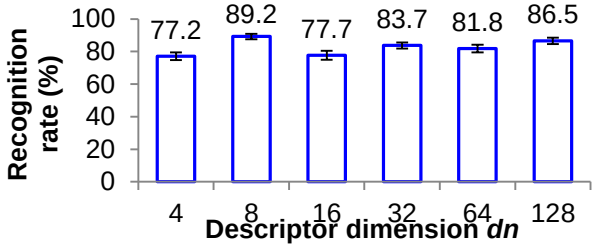


Fig. 7. Recognition rates with different descriptor dimensions *dn*.

As in general the manipulations in hand are swift, they have a strict requirement for time efficiency. Because the classification procedures with same dictionary size and touch times take the same time, the main difference happens for different descriptors is the time taken to compute descriptors from the raw tactile images. The time taken to process one tactile image to acquire 3 descriptors of different dimensions is listed in Table 1. It can be observed that it takes more time to compute the descriptors with 8 directions than those with 4 directions and for both two cases, descriptors of lower dimensions cost less time.

Table 1 Processing time of one tactile image to obtain 3 descriptors

dn	4	8	16	32	64	128
Time/ms	1.9	2.2	2.0	2.4	2.1	2.6

Taking both the classification performance and time efficiency into account, 8-dimensional descriptors are proposed to be used as Tactile-SIFT features in tactile recognition.

2) Number of sub-patches sp

In this case, *sp* is investigated and varied whereas *dn*, *k* and *m* are fixed. 8-dim Tactile-SIFT descriptors are used (*dn*=8); dictionary size *k* is set to 50; the number of touches *m* for one test trial is chosen as 15. In a stepwise manner, one patch is divided into 1 (no segmentation), 2, 3, 4 and 5 sub-patches. For the cases of 1 and 2 sub-patches, the tactile readings are resampled but for the cases of 4 and 5 sub-patches, the tactile readings are interpolated along the vertical axis. Inspired by [25], the size of all these sub-patches is set to 6×6 and there are overlaps of 3 along vertical axis between the adjacent sub-patches. As shown in Figure 8, the classification accuracy has a dramatic decrease when the tactile reading is resampled due to information loss. It maintains a similar performance as the case of 3 sub-patches when the images are interpolated; in other words, adding virtual information has slight effect on the accuracy. But increasing partition brings more time burden to the system, thus it is chosen to divide each reading into 3 sub-patches.

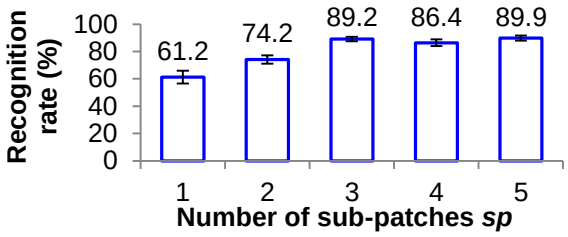


Fig. 8. Overall classification accuracies with different number of sub-patches.

3) Dictionary size k

In this case, *k* is investigated and varied whereas *dn*, *sp* and *m* are fixed as follows. 8-dim descriptors are used (*dn*=8); one tactile image is segmented into three sub-patches (*sp*=3); *m*=10, 12 and 14 touches are utilized. It is apparent that the larger the dictionary size *k* is the higher recognition rates can be achieved. As shown in Figure 9, 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" happen when the size increases more, like when we describe objects. Thus at last a dictionary size of 50 is chosen.

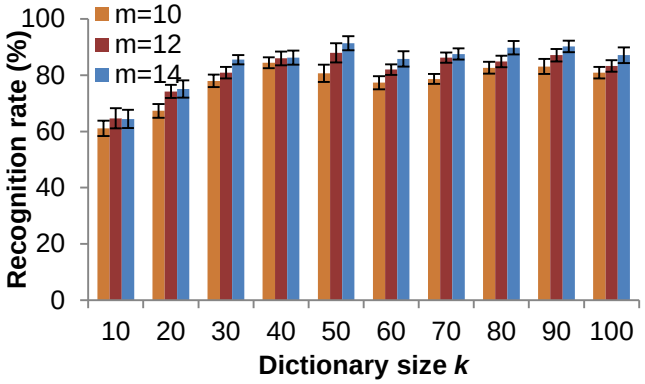


Fig. 9. Comparison of overall accuracies with various dictionary sizes.

4) Number of touches m

Besides, the effect of the number of touches *m* on the recognition rates is also investigated. In this case, *dn*, *sp* and *k* are fixed as follows. 8-dim descriptors are used (*dn*=8); one tactile image is segmented into three sub-patches (*sp*=3); *k*=40, 50 and 60 are used for comparison. As for the evaluation of dictionary size *k*, the classification procedure (25 test trials each) is carried out 10 times and the mean values are calculated to mitigate the uncertainties. It is shown in Figure 10 that the more times the robot touches objects the more reliable it recognizes them correctly. But reasonable accuracy could be obtained when 15 samples are collected; around 90 percent objects can be recognized correctly. Hence in conclusion the robot only needs a few observations to reach a reasonable guess.

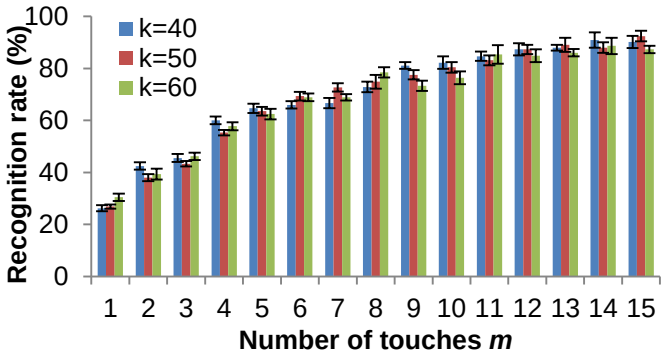


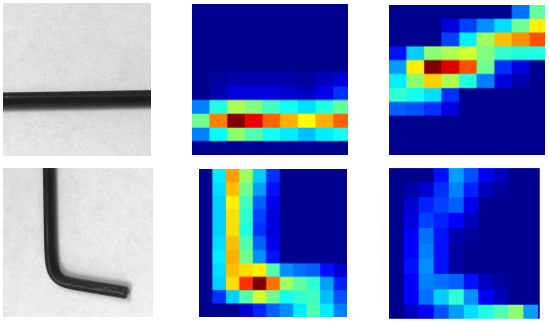
Fig. 10. Comparison of overall accuracies with different number of samples.

5) Classification results

Based on the discussion, in our final test, 8-dim Tactile-SIFT descriptors are used (*dn*=8); one tactile image is segmented into three sub-patches (*sp*=3); the dictionary size *k* is set as 50 and *m*=15 samples are used for each test trial. As a result, an overall classification accuracy of 85.46% is achieved by averaging the cross validation results and the confusion matrix with a recognition rate of 91.33% is shown in Figure 11. It proves the robustness of our algorithm w.r.t. different poses and relative positions between objects and the tactile sensor. On the other hand, some of the objects are assigned to wrong labels, e.g., some observations of the plier and wrench are wrongly concluded to be from the fixed wrench, which is caused by their common features such as lines.

The robustness w.r.t. pose variance is also validated. The readings from a line shape orientated at an angle from 0 to 360° at an internal of 10°, as shown in Figure 12, are assigned to the same codeword with a rate of 88.9%. And the readings from a

corner shape orientated at the same orientation range are assigned to the same codeword with a rate of 81.5%.



(a) (b) (c)

Fig. 12. Images of a line shape (upper) and a corner (lower) (a) and tactile readings obtained when they are put at 0^o (b) and 30^o (c).

V. CONCLUSION AND DISCUSSION

In this paper it is proposed to recognise objects invariant to their movements with a tactile sensor by using a novel local Tactile-SIFT feature in a framework of bag of words. The SIFT descriptors are adapted to process tactile images to extract features from segmented tactile sub-patches and different dimensional Tactile-SIFT descriptors are evaluated and compared. It is found that the 8-dimensional descriptors outstand taking both classification accuracy and time efficiency into consideration. The methods to segment tactile images are also tested. Based on the acquired descriptors, a vocabulary of *k* words is learned by *k*-means unsupervised learning and the histogram codebook is used to identify objects by *k*NN. The proposed system is validated with an off-the-shelf tactile sensor and high classification accuracy 91.33% has been achieved.

The proposed method can be extended in the following aspects. 1) The classification accuracy in our experiments tends to reach a limit of around 90%. One of appropriate reasons is that the locations of the features obtained on the object are not considered in the current method. However, intuitively, these two information sources are correlated. Therefore, we will look into incorporating such information to improve the classification accuracy. 2) In this paper most investigated shapes are in 2D; thus in the future work it will be extended to the 3D object recognition. 3) It is also expected to extend the algorithm to recognizing contacts with multiple tactile sensing pads, such as an instrumented robotic hand with multiple tactile array sensors on the fingers and the palm. We will explore how to recognize the object with fewer touches in such case.

REFERENCES

[1] S. J. Lederman and R. L. Klatzky, "Haptic perception : A tutorial," *Attention, Perception,& Psychophys.*, vol. 71, no. 7, pp. 1439–1459, 2009.

[2] R. S. Dahiya, P. Mittendorfer, M. Valle, G. Cheng, and V. J. Lumelsky, "Directions Toward Effective Utilization of Tactile Skin : A Review," *IEEE Sens. J.*, vol. 13, no. 11, pp. 4121–4138, 2013.

[3] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta, "Methods and Technologies for the Implementation of Large-Scale Robot Tactile Sensors," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 389–400. 2011.

[4] M. Kaltenbrunner, T. Sekitani, J. Reeder, T. Yokota, K. Kuribara, T. Tokuhara, M. Drack, R. Schwödiauer, I. Graz, S. Bauer-Gogonea, S. Bauer, and T. Someya, "An ultra-lightweight design for imperceptible plastic electronics.," *Nature*, vol. 499, no. 7459, pp. 458–63, 2013.

[5] H. Xie, H. Liu, S. Luo, L. D. Seneviratne, and K. Althoefer, "Fiber optics tactile array probe for tissue palpation during minimally invasive surgery," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 2539–2544.

[6] H. Wu, H. Liu, and D. Liu. "Two-Dimensional Direction Recognition Using Uniaxial Tactile Arrays," *IEEE Sens. J.*, vol. 13, no. 12, pp. 4897-4903, 2013.

[7] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile Sensing—From Humans to Humanoids," *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 1–20. 2010.

[8] N. Jamali and C. Sammut, "Majority Voting : Material Classification by Tactile Sensing Using Surface Texture," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 508–52, 2011.

[9] S. Decherchi, P. Gastaldo, R. S. Dahiya, M. Valle, and R. Zunino, "Tactile-Data Classification of Contact Materials Using Computational Intelligence," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 635–639, 2011.

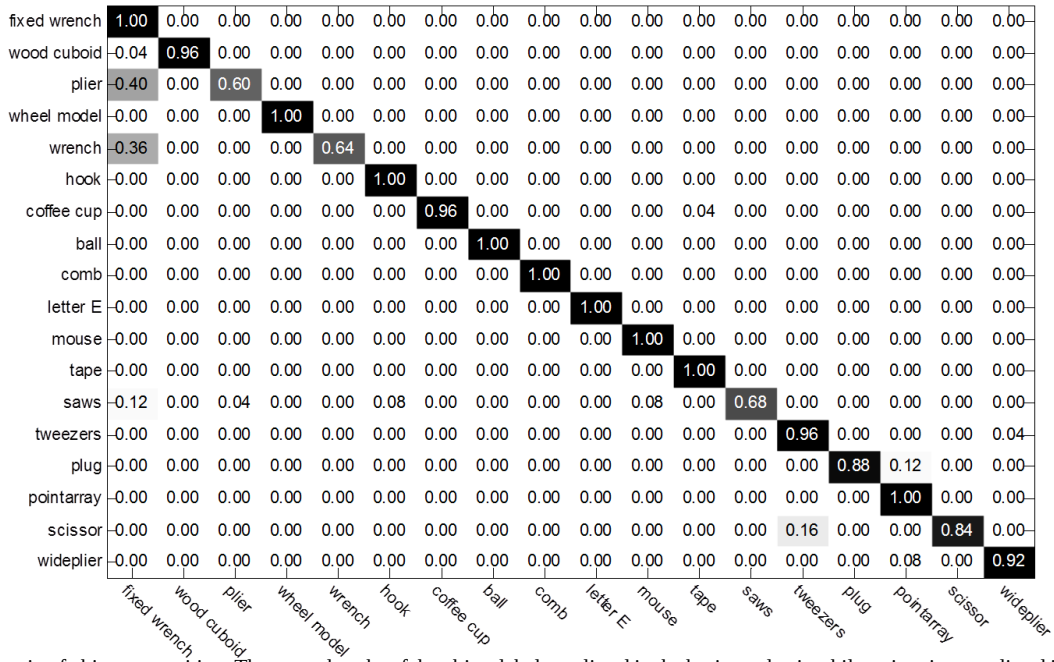
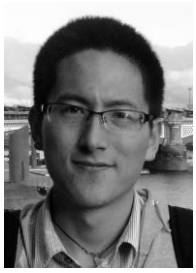


Fig. 11. Confusion matrix of object recognition. The ground truths of the object labels are listed in the horizontal axis while estimations are listed in the vertical axis.

[10] H. Liu, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer, "Surface material recognition through haptic exploration using an intelligent contact sensing finger," in *International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 52–57.
[11] R. S. Fearing and T. Binford, "Using a Cylindrical Tactile Sensor for Determining Curvature," *IEEE Trans. Robot. Autom.*, vol. 7, no. 6, pp. 806–817, 1991.
[12] Y. Jia and J. Tian, "Surface Patch Reconstruction From 'One-Dimensional' Tactile Data," *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 2, pp. 400–407. 2010.
[13] R. A. Russell and S. Parkinson, "Sensing surface shape by touch," in *IEEE Conference on Robotics and Automation (ICRA)*, 1993, pp. 423–428.
[14] P. K. Allen and K. S. Roberts, "Haptic object recognition using a multi-fingered dextrous hand," in *IEEE International Conference on Robotics and Automation (ICRA)*, 1989, pp. 342 – 347 vol.1.
[15] P. K. Allen and P. Michelman, "Acquisition and interpretation of 3-D sensor data from touch," *IEEE Trans. Robot. Autom.*, vol. 6, no. 4, pp. 397–404, 1990.
[16] M. Charlebois, "Shape Description of Curved Surfaces from Contact Sensing Using Surface Normals," *Int. J. Rob. Res.*, vol. 18, no. 8, pp. 779–787, 1999.
[17] M. Charlebois, K. Gupta, and S. Payandeh, "Shape Description of General, Curved Surfaces Using Tactile Sensing and Surface Normal Information," in *IEEE International Conference on Robotics and Automation (ICRA)*, 1997, pp. 2819–2824.
[18] 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 *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1410–1415.
[19] H. Liu, J. Greco, X. Song, J. Bimbo, and K. Althoefer, "Tactile Image Based Contact Shape Recognition Using Neural Network," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2013, pp. 138–143.
[20] A. R. Jimenez, A. S. Soembagijo, D. Reynaerts, H. Van Brussel, R. Ceres, and J. L. Pons, "Featureless classification of tactile contacts in a gripper using neural networks," *Sensors Actuators A Phys.*, vol. 62, pp. 488–491, 1997.
[21] Z. Ji, F. Amirabdollahian, D. Polani, and K. Dautenhahn, "Histogram based classification of tactile patterns on periodically distributed skin sensors for a humanoid robot," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2011, pp. 433–440.
[22] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 243–248.
[23] A. Khasnobish, G. Singh, A. Jati, A. Konar, and D. N. Tibarewala, "Object-shape recognition and 3D reconstruction from tactile sensor images," *Med. Biol. Eng. Comput.*, vol. 52, no. 4, pp. 353–362. 2014.
[24] 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.
[25] Z. Pezzementi, E. Plaku, C. Reyda, and G. D. Hager, "Tactile-Object Recognition From Appearance Information," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 473–487, 2011.
[26] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[27] Z. Pezzementi, C. Reyda, and G. D. Hager, "Object mapping, recognition, and localization from tactile geometry," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 5942–5948.
[28] M. Meier, M. Schopfer, R. Haschke, and H. Ritter. "A probabilistic approach to tactile shape reconstruction." *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 630-635, 2011.
[29] E. Nowak, F. Jurie, and B. Triggs, "Sampling Strategies for Bag-of-Features," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 490–503.
[30] S. Luo, W. Mou, M. Li, L. D. Seneviratne, K. Althoefer, and H. Liu, "Rotation and Transposition Invariant Object Recognition with a Tactile Sensor," in *IEEE Sensors Conference*, 2014, pp. 1030–1033.
[31] A. Bosch, A. Zisserman, and X. Mu, "Scene Classification via pLSA," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 517–530.
[32] F.-D. Jou, K.-C. Fan, and Y.-L. Chang, "Efficient matching of large-size histograms," *Pattern Recognit. Lett.*, vol. 25, no. 3, pp. 277–286, 2004.



Shan Luo received the B.S. degree in engineering from China University of Petroleum, Qingdao, China, in 2012. With funding from the King's-China Scholarship Council (K-CSC) he is currently pursuing the Ph.D. degree at the Centre for Robotics Research, King's College London. His research interests include tactile sensing, object recognition and computer vision.

Wenxuan Mou received the B.S. degree in Automation from China University of Petroleum, China, in 2012 and the M.S. degree in Computer Vision from Queen Mary University of London, UK, in 2014. She is currently pursuing the Ph.D. degree in Computer Vision at Queen Mary University of London. The topic of her research is Multimodal analysis of affective and social dimensions of group behavior. Her research interests include affective computing, human-computer interaction and machine learning.

Kaspar Althoefer (M'03) received his Dipl.-Ing. degree in electronic engineering from the University of Aachen, Aachen, Germany, and his Ph.D. degree in electronic engineering from King's College London, London, U.K. He is currently a Professor of Robotics and Intelligent Systems and Head of the Centre for Robotics Research (CoRe), Department of Informatics, King's College London. He has been involved in research on mechatronics since 1992 and gained considerable expertise in the areas of sensing, sensor signal analysis, embedded intelligence, and sensor data interpretation using neural networks and fuzzy logic, as well as robot-based applications He has authored or coauthored more than 200 refereed research papers related to mechatronics and robotics.

Hongbin Liu is currently a lecturer (Assistant Professor) in the department of Informatics, King's College London, UK. He received his B.S. degree in 2005 from the Northwestern Polytechnique University, Xi'an, China, and was awarded with an MSc degree from King's College London in 2006. He received his PhD in 2010 from Kings College London. He is a member of IEEE. His research interests include tactile/force perception based robotic cognition, modeling of dynamic interaction, medical robotics and haptics.