A Visuo-Haptic Framework for Object Recognition Inspired from Human Tactile Perception †

Ghazal Rouhafzay [1], and Ana-Maria Cretu [2]

[1] Department of Systems and Computer Engineering, Carleton University; ghazal.rouhafzay@carleton.ca
[2] Department of Systems and Computer Engineering, Carleton University; anamaria.cretu@carleton.ca
* Correspondence: ghazal.rouhafzay@carleton.ca;
† Presented at the 5th International Electronic Conference on Sensors and Applications, 15–30 November 2018; Available online: https://sciforum.net/conference/ecsa-5.

Abstract: This paper addresses the issue of robotic haptic exploration of 3D objects using an enhanced model of visual attention, where the latter is applied to obtain a sequence of eye fixations on the surface of objects guiding the haptic exploratory procedure. According to psychological studies, somatosensory data resulted as a response to surface changes sensed by human skin is used in combination with kinesthetic cues from muscles and tendons to recognize objects. Drawing inspiration from these findings, a series of 5 sequential tactile images are obtained by adaptively changing the size of the sensor surface according to the object geometry for each object, from various viewpoints, during an exploration process. We take advantage of the contourlet transform to extract several features from each tactile image. In addition to these somatosensory features, other kinesthetic inputs including the probing locations and the angle of the sensor surface with respect to the object in consecutive contacts are added as features. The dimensionality of the large feature vector is then reduced using a self-organizing map. Overall 12 features from each sequence are concatenated and used for classification. The proposed framework is applied to a set of four virtual objects and a virtual force sensing resistor array (FSR) is used to capture tactile (haptic) imprints. Trained classifiers are tested to recognize data from new objects belonging to the same categories. Support vector machines yield the highest accuracy of 93.4%.

Keywords: Haptic exploration; visual attention; visuo-haptic interaction; tactile object recognition

1. Introduction

Many psychological research articles over the past three decades focus on the haptic perception and exploratory procedures employed by humans to identify objects and their characteristics. Lederman et al [1] identify six manual exploratory procedures exploited by humans when interacting with an object, among which "enclosure" and "contour following" provide information about the global and exact shape of objects respectively. The authors also refer to two different types of information being used during these exploratory procedures: from mechanoreceptors in the skin (cutaneous cues), capturing fine textural details, and from mechanoreceptors in joints and tendons (kinesthetic cues) for geometrical shape identification. Reproducing such haptic exploration techniques for humanoid robots has recently attracted a wide research interest. On the other hand, tactile perception is proven to be more reliable in presence of vision [2], and these two senses contribute closely in human sensorial loop. Inspired from the visuo-haptic interaction in human sensorial loop and the sequential nature of haptic exploration to integrate several tactile features of objects, in this work we have developed a framework for robotic object recognition based on tactile probing from a sequence of eye fixations on the object surface. Accordingly, an enhanced model of visual attention is employed to determine a sequence of eye fixations on the surface of objects from different viewpoints. Subsequently tactile data are collected by adaptively changing the sensor

surface size according to the local object geometry. These cutaneous cues together with the 3D coordinates of probing locations and the normal vector of object surface at probing locations as two kinesthetic cues are used for object recognition. To confirm the efficiency of our framework before implementation, we have performed experiments using a virtual tactile sensor and virtual 3D models. The paper is organized as follows: A brief literature review of related work on tactile object recognition is provided in section 2. Section 3 presents the proposed framework. Obtaining the sequence of eye fixations and the adaptive collection of tactile data are discussed in section 3.1 and 3.2, respectively. Section 3.3 discusses feature extraction, while section 3.4 presents the kinesthetic cues employed in object recognition. The data processing strategy is introduced in section 3.5. Finally, section 4 shows obtained results and section 5 concludes the work.

2. State of the art

Enabling humanoid robots with a capability of sensing their environment similar to humans is a challenging subject in recent robotic research. Artificial sense of touch allowing the identification of a wide range of object characteristics including the deformability, elasticity, textural features, temperature, approximate weight, etc. is a beneficial technology attracting a huge research interest. Accordingly, a variety of tactile sensors are designed and manufactured. In a recent publication, Chi et al. [3] discuss the latest advancements in technology of tactile sensors. Tactile arrays developed using force sensing resistors have demonstrated a high reliability for object recognition. Liu et al. [4] use a three-finger robot to collect tactile sequences. Dynamic time wrapping between sequences is then computed to measure dissimilarities for classification based on joint sparse coding. Authors in [5] capture tactile data as displacement of finger joints of a robot when grasping an object. They train a self-organizing map (SOM) for classification purposes. Luo et al. [6] employ the 3D coordinates of probing locations with the index of clustered tactile data using k-mean algorithm for object classification. Gorges et al. [7] take benefit from a robotic hand with five fingers to recognize a set of seven objects based on haptic data acquired from a sequence of palpations. Literature has also witnessed several attempts in integration of haptic and visual data for object recognition. Gao et al. [8] train a deep neural network architecture for object classification by learning both haptic and visual features. They demonstrate that the integration of visual and haptic features outperforms the case where visual and tactile characteristics are employed separately. In our previous work [9], we developed a framework for object recognition where tactile data are gathered from visually salient regions with the aim to overcome the high computational cost of probing the whole surface of objects. In this work we take advantage of a model of visual attention to comply to the sequential nature of haptic exploration with a sequence of eye fixations. To reproduce the exploration strategies performed by humans where the global shape of object is perceived in general by the palm and finer details are captured by fingertips, the size and precision of the tactile sensor adaptively changes based on the geometrical characteristics of objects.

3. Framework

Figure 1 illustrates our framework for object recognition based on guided haptic exploration. The first two stages of the work correspond to the development of a model of visual attention and to the determination of the sequential strategy to move the tactile sensor. For this purpose, the virtual camera of Matlab turns around each object to capture images from 16 viewpoints and a computational model of visual attention [10] is used to determine the sequence of eye fixations from each viewpoint. The tactile sensor then follows the sequences of eye fixations to collect tactile imprints at their locations. In next stage, a vector of 16 features is computed for each tactile image. The 3D normal vector and the 3D coordinate of probing locations add up six other tactile features. Consequently, five vectors (as we use series of 5 sequential tactile images) of size 1×22 are computed for each sequence of eye fixations. To reduce the dimensionality of feature vectors, a self-organizing map (SOM) is trained, resulting in five-dimensional feature vectors. The standard deviation, rms value, and skewness of each sequence and the same measures extracted from wavelet coefficients for a 3-level decomposition of sequences by *daubechies* 2 wavelets are concatenated and

fed to five classifiers, namely k-nearest neighbors (kNN), support vector machine (SVM), decision trees, quadratic discrimination and Naïve Bayes for classification. In this work, we have conducted experiments over four classes of objects each containing three objects, two of which are used for training the classifiers and the third one for testing. Further details are provided in following sections.
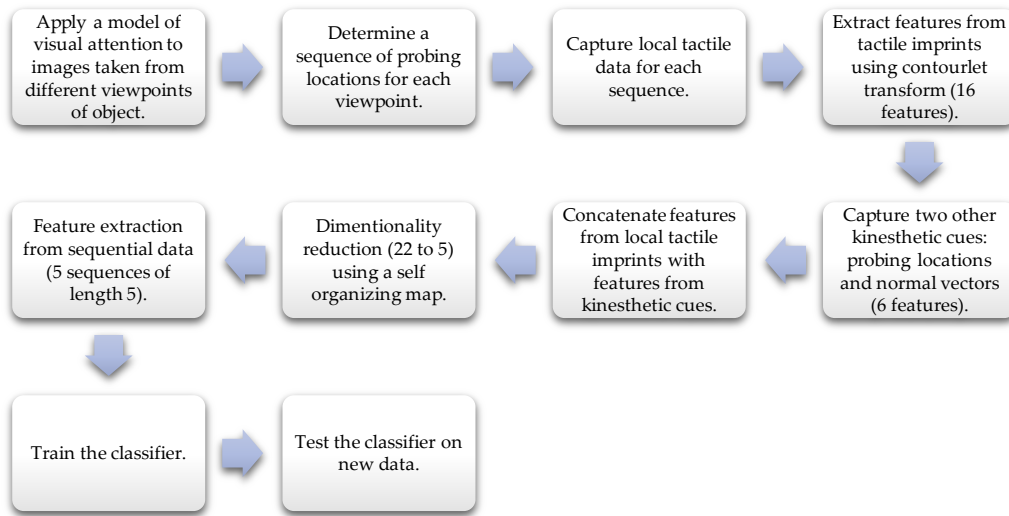

Figure 1. Framework of the introduced tactile object recognition strategy.

3.1 Sequences of eye fixations

In human visual system, the allocation of the narrow high-resolution part of retina (fovea) permitting the full visual perception is referred to as focus of attention. Research from neuroscience confirms the contribution of a series of features extracted in field of view such as color opponency, contrast, curvature, intensity, orientation etc. in the allocation of attentional resources. Accordingly, researchers tried to reproduce this process as a computational model of visual attention for a rapid automatic analysis of scenes. In this work, we have adopted the enhanced model of visual attention presented in [10] to compute saliency maps for 3D objects. These saliency maps assign higher intensities to regions attracting the attention based on which a sequences of eye fixations can be retrieved for each viewpoint of the object in order of importance. Only the first 5 elements of the sequence of eye fixations for each object are taken into consideration in the remainder of the paper. The tactile sensor follows this sequence of eye fixations to collect tactile data in view of object classification. Figure 2 summarizes the procedure of obtaining the sequence of eye fixations.
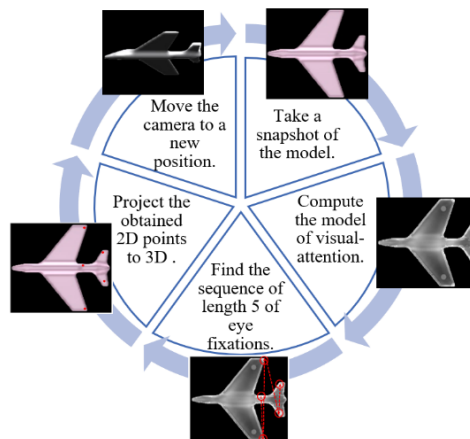

Figure 2. Determination of the sequence of eye fixations for each viewpoint.

3.2 Tactile data collection using adaptive sensor size

In force sensing resistors (FSR)-based tactile sensors, the deformation on the surface of the elastic surface covering the FSR array, when subjected to an external force and in contact with the surface of an object, is transduced to produce a tactile image. In this work, a virtual tactile sensor is simulated such that the deformation in sensor's surface is measured as the distance between points on a tangential plane (representing the surface of the sensor) and the object, when the center of the plane is in direct contact with the object. The obtained values are then normalized between zero and one to yield a tactile image. Using the probing locations at the determined eye fixations in section 3.1, the center of sensor array is considered positioned at these locations and simulated to touch the probing point. As a result, in certain situations such as concave surfaces, negative distances between the object and sensor surface can be achieved indicating an intersection between the sensor and object. Figure 3 illustrates an example of such a case. Since in such probing cases data cannot be acquired in reality (i.e. using real rigid backing FSRs), and in compliance with the haptic exploration strategy by humans where the overall shape of objects is probed by the palm (large tactile surface) and finer details can be captured by fingertips (small tactile surface), in current research the sensor's surface is adaptively diminished to capture the local tactile data. However, to keep the size of tactile image consistent (i.e. 32×32 during experimentation), the distance between the sensing points is reduced, thus resulting in a higher local precision.

Figure 3. Adaptive modification of tactile sensor size.

3.3. Feature extraction from tactile imprints

Feature extraction from tactile data is a determining factor in object recognition. Authors in [11] prove the efficiency of wavelet decomposition in feature extraction from tactile data. On the other hand, contourlet transformation is believed to outperform the wavelet decomposition to extract features when applied to images [12]. Consequently, in this paper we have used the contourlet transformation to extract features from tactile imprints. As such, a 16 directional contourlet transform [12] is first applied to each tactile image and then the standard deviation of obtained coefficients for each directional sub-band is computed to produce a feature vector of size 16.

3.4. Kinesthetic cues

Human skin is not the only source of tactile information. When exploring an object with the hand, the angle between finger phalanges and the position of fingers as they are in contact with the object surface, supply crucial information about the size and shape of the explored objects, information that is not available from the skin mechanoreceptors. In neuroscience such data from joints, bones and macules are referred to as kinesthetic cues. In this paper to employ kinesthetic cues for object recognition, we have added the 3D coordinates of the probing locations as well as the local normal vectors at the probing points as features for classification.

3.5 Data processing

As previously explained, we use 16 features extracted from each tactile imprint in addition to the 3D coordinates and the normal vector of probing points, thus resulting in a final feature vector of size 1×22. Five consecutive imprints captured over the sequence of eye fixations are used for classification. As high dimensional data has a negative impact on classification accuracy, we train a self-organizing map to reduce the 22-dimensional feature vector to only five dimensions, thus

resulting in five sequences (as we have five features) of length five (five tactile data from the sequence of eye fixations). To train a classifier, we need to determine how each feature varies over the sequence when moving the tactile sensor. For this purpose, we take advantage of *daubechies* 2 wavelet to decompose each sequence into 3 levels. The standard deviation, rms value and skewness of the wavelet coefficients for each level as well as those of the sequence itself are concatenated to produce a final feature vector for classification.

4. Experiments and results

Figure 5 illustrates the 3D models used for experimentation. The objects in first two columns are used to train classifiers and the tactile data from objects in the third column are used for testing. As briefly mentioned before, five classifiers, namely kNN, SVM, decision trees, quadratic discrimination and Naïve Bayes are trained using the generated data set according to the previously explained data processing strategy.
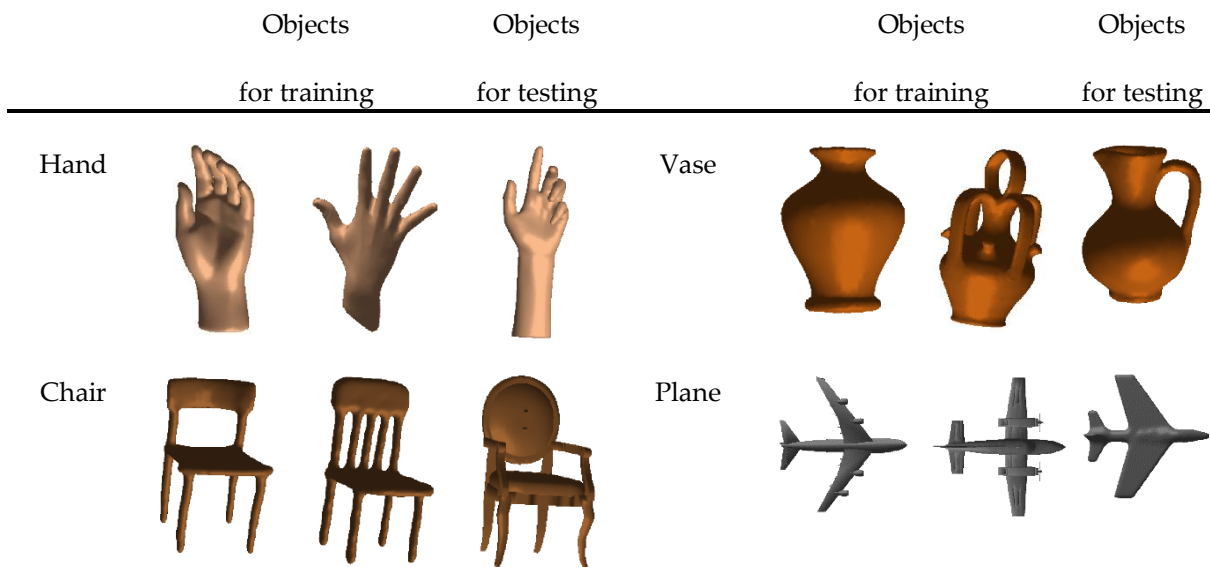

Figure 4. 3D models used for training and testing.

The results obtained using the proposed framework are compared in Table 1 with the case where sequential data are captured by randomly moving the sensor over the object surface. When collecting data at random positions over the surface of the object, we also use a sequence of 5 tactile imprints for each object. As one can notice in this table, results confirm that in most cases, the performance when using the sequence of eye fixations to guide the probing outperforms the case where tactile data is acquired at random positions over the surface of objects. SVM demonstrates a superior performance for sequence classifications compared to other classifiers. Quadratic discrimination with a maximum difference of 1.86% closely competes with the SVM.

Table 1. Object classification results

Classifier	Guiding the tactile sensor by a sequence of 5 eye fixations	Using a random movement of the tactile sensor
kNN	85.58%	84.94%
SVM	**93.45%**	86.14%
Decision trees	73.56%	71.63%
Quadratic discrimination	91.89%	84.28%
Naïve Bayes	56.68%	57.16%

5. Conclusion

In this work we have proposed a framework for tactile object recognition where a model of visual attention is used to guide sequential tactile exploration. A virtual tactile sensor is simulated to collect tactile data. Inspired from human tactile exploration and object recognition the size of tactile sensor adaptively modifies to capture tactile images. We have employed contourlet transformation to extract features from tactile imprints. Two kinesthetic cues as the normal vectors and 3D coordinates of probing locations are added to provide information about the size and general shape of objects. A self-organizing map is trained to reduce the high dimensional features. A series of features from sequences are then extracted to construct a final data set. Five different classifiers are trained and tested over data from a new object. Support vector machines and quadratic discrimination classifiers with accuracies of 93.45% and 91.89% respectively achieve the highest accuracies. We thus demonstrate that employing the sequence of eye fixations to guide the tactile probing operation enhances the classification accuracies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lederman, S. J.; Klatzky, R. Haptic perception: A tutorial, *Attention, Perception, & Psychophysics*, **2009**, *Volume 71*, no. 7, pp. 1439-1459.
2. Klatzky, R. L.; Lederman S. J.; Matula, D. E. Haptic Exploration in the Presence of Vision, *Human Perception and Performance*, **1993**, *Volume 19*, no. 4, pp. 726-743.
3. Chi, C.; Sun, X.; Xue, N.; Li, T.;Liu, C. Recent Progress in Technologies for Tactile Sensors, *Sensors* (Basel, Switzerland), **2018**, *Volume 18*, no. 4, 948.
4. Liu, H.; Guo, D.; Sun, F. Object Recognition Using Tactile Measurements: Kernel Sparse Coding Methods, *IEEE Trans. on Instrumentation and Measurement*, **2016**, *Volume 65*, no. 3, pp. 656-665.
5. Ratnasingam, S. and McGinnity, T. Object recognition based on tactile form perception, *IEEE Workshop Robotic Intelligence in Informationally Structured Space*,**2011**, Paris.
6. Luo, S.; Mou, W.; Althoefer, K.; and Liu, H.   iCLAP: shape recognition by combining proprioception and touch sensing, *Autonomous Robots*, **2018**, pp. 1-12.
7. Gorges, N.; Navarro, S. E.; Goger, D.; Worn, H. Haptic Object Recognition using Passive Joints and Haptic Key Features, *IEEE International Conference on Robotics and Automation*, **2010**, Anchorage, Alaska, USA.
8. Gao, Y.; Hendricks, L.; Kuchenbecker, K. J. Deep learning for tactile understanding from visual and haptic data, *IEEE International Conference on Robotics and Automation (ICRA)*, **2016**, Stockholm.
9. Rouhafzay, G.; Pedneault, N.; Cretu, A.-M.   A 3D Visual Attention Model to Guide Tactile Data Acquisition for Object Recognition, in *Proceedings of the 4th International Electronic Conference on Sensors and Applications*, **2017**.
10. Rouhafzay G.; Cretu, A.-M.   Perceptually Improved 3D Object Representation Based on Guided Adaptive Weighting of Feature Channels of a Visual-Attention Model, *3D Research*, 2018, *Volume 9*, no. 3.
11. Adi, W. and Sulaiman, S. Using Wavelet Extraction for Haptic Texture Classification, Visual Informatics: Bridging Research and Practice. IVIC, *Lecture Notes in Computer Science*. **2009**, *Volume 5857*, pp. 314-325.
12. Do, M. N. and Vetterli, M. The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, *IEEE Trans. Image Processing*, **2005**, *Volume 14*, no. 12, pp. 2091-2106.