Towards Reinforcement Learning of Haptic Search in 3D Environment

Alexandra Moringen[1], Julian Nowainski[2], and Helge Ritter

CITEC, Bielefeld University, Inspiration 1, 33619, Germany
abarch@techfak.uni-bielefeld.de,
WWW home page:
https://ni.www.techfak.uni-bielefeld.de/people/abarch

Abstract. Due to a high computational cost, reinforcement learning in robotic applications is often based on non-sparse reward functions. We believe that in order to enable a multi-fingered anthropomorphic robot to autonomously perform haptic search in 3D environment, a reward function enforcing accurate classification between a target and a distractor is essential. To this end, we investigate performance of a target-distractor classifier that employs Generalized Matrix Learning Vector Quantization (GMLVQ). This method is particularly suitable for this task as it is designed to represent one class with several prototypes. This fits well to one of our main requirement of representing haptic exploration of a range of geometric features under one label. Apart from a suitable approach to classification, GMLVQ illustrates relevances of input dimensions, and can be used for data visualization in low-dimensional space.

Keywords: artificial touch, haptic search, relevance learning, target-distractor classification

1 Introduction

Sense of touch, or haptic perception is the most robust and ubiquitous human sense. Haptic interaction with the objects that surround us, involving both haptic attention and haptic memory, contributes to the success of a prevailing part of our daily manual tasks. Coupled with the haptic memory and attention, proprioception and haptic perception in particular enables us to seamlessly perform tasks that have not been yet implemented even on the most advanced robotic platforms. This is due to the fact that observation, understanding and modeling of haptic interaction in a three-dimensional environment has until now remained sparse. Implementation of a haptics-based manual control to endorse autonomous haptic interaction for anthropomorphic robots in a three-dimensional environment is an open research question. We discuss our approach to this question in the following points. Firstly, Modular Haptic Stimulus Board (MHSB) provides extensive possibilities to observe different aspects of haptic interaction in a wide range of scenarios and different degrees of complexity. Human data recorded with a set of high performance hardware that enable tactile, kinematic, and joint angle recording during haptic interaction, provides a foundation for an extensive modeling approach [5]. Figure 1 illustrates the most recent experimental setup containing two MHSBs.

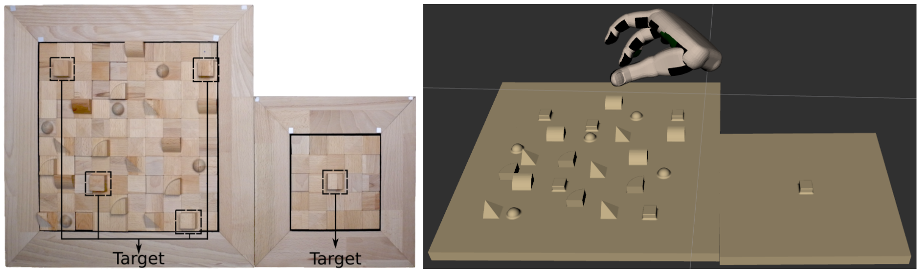

Fig. 1. Exemplary stimulus material used in the experimental setup (left), visualization of the setup and the acquired data in RViz (right). A full trial visualization including the acquired Vicon, tactile and joint angles data is available under the link https://www.techfak.uni-bielefeld.de/persons/abarch/td_viz.ogv. In a trial blindfolded participants were asked to memorize a target located in the smaller board, and to find it in the larger board by using touch only.

Secondly, reinforcement learning is undoubtedly one of the most promising frameworks to approach the above described research question. However, in order to quickly and successfully find an optimal policy, a suitable reward function is essential. We believe that the key is to define the initial non-sparse reward function based on a target-distractor classifier certainty. Such a classifier has to be initially trained on an equivalent data acquired by a human - a valid policy to generate such data is not available in the beginning. By enforcing a policy that results in a high object recognition certainty, an initial network training is likely to generate a meaningful behavior. In this paper we briefly discuss classification results and the corresponding relevance values based on GMLVQ [7]. A detailed description of the experiment can be found in [6, 5].

2 METHODS

2.1 Stimulus Material and Experimental Scenario

The stimulus material of the MHSB has been previously employed in a range of studies [3, 2, 4], and represents a three-dimensional shape environment through a combination of wooden bricks. In this study, five identical object copies per object class $l \in \{1,\dots,5\}$ have been used: wave, cylinder, cuboid, sphere, and pyramid (see [5]). Altogether 25 bricks with shapes carved on top have been used. In turns, one of the object classes l has been employed as a target and the other object classes $\{1,\dots,5\} \setminus \{l\}$ as distractors. On the right 5×5-brick board (see Figure 1 - left), only one object serving in the role of the target has been presented to the study participants. The rest of the board has been filled with planar-surfaced neutral bricks. The left 10×10-brick board illustrated in the same figure represented the search environment with all remaining 24 bricks of all five object classes, presenting four occurrences of the target object randomly distributed among four distractor object classes and the neutral bricks.

2.2 Task and Procedure

The study participants were blindfolded and performed the following three-staged task in the experimental setting previously discussed in Section 2.1. A detailed description of the task can be found in [6, 5]:

1. **Memorize the target object** presented in the small MHSB on the right.
2. **Search** for multiple instances of the target object in the large MHSB on the left (the search environment). Perform the task as fast as possible, and memorize as many positions of the target objects as possible, until the time limit is reached.
3. **Verify the success of the performed search** by going back to the search environment and retrieving the target object placement from memory.

2.3 Classification approach

In this section we provide only a brief description of the GMLVQ and refer to [7] due to space restrictions. The foundation of the GMLVQ is the LVQ [1], a supervised classification scheme that represents classes by prototypes. GMLVQ is a metric learning scheme for a LVQ classifier that distorts the Euclidean distance between classes to improve classification by making similar points more similar and dissimilar points more dissimilar. The method extends the parameter space with the so-called *relevance matrix* Λ resulting in the following distance measure: $d^{\Lambda}(\boldsymbol{w}, \boldsymbol{\xi}) = (\boldsymbol{\xi} - \boldsymbol{w})^T \Lambda (\boldsymbol{\xi} - \boldsymbol{w})$, where $\boldsymbol{\xi}, \boldsymbol{w} \in \mathbb{R}^N$ denotes a data point and a prototype, respectively.

In our work we have trained five target-specific classifier (see [5] for details). Each one of the five classifiers has been trained to assign one of the three labels: exploration of the target, exploration of a distractor, hand in the air. The number of points per class was approximately 9000.

3 Results and Discussion

In our preliminary evaluations of GMLVQ classification we have received accuracy of approx. 70-80% for the complete data, and accuracy of 80-90% by evaluating only 30% of the points from the end of each segment.[1] The errors are very likely to be caused by similarities between the objects. Therefore, a point-wise classification certainty may be used as a component of a reward function. Importantly, we have already seen an improvement of both classification certainty and a decrease in error rate towards the end of the segment in our previous work [5]. We believe that a recurrent temporal accumulation of either the resulting classification certainty or the features extracted from the data should be performed in future work. This may guide the decision-making processes w.r.t. continuing to explore the current object, or switching to a new one.

An example of the relevance values calculated for a target-specific classifier (see Figure 2) illustrates that the tactile fingertip sensors (endings *do, *dm and *di) are

[1] Segments are formed by uniting all subsequent time series points with a given object label, until the label changes. Once a participant switches haptic exploration to a different object, the label changes, and a segment border is created.

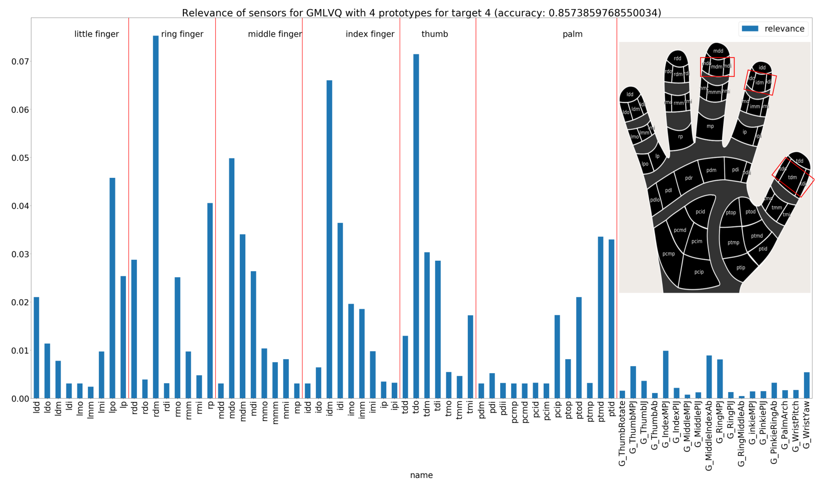

Fig. 2. Relevance values for all available tactile and joint angle sensor values. Middle finger, index finger and thumb demonstrate similar patterns of the relevance values.

strongly relevant. For joint angles, the metacarpal joints (ending *MPJ) and the abduction between the index and the middle finger are relevant for this particular classifier. However, the difference between the target-specific classifiers is large w.r.t. the distribution of relevance values within a given individual above-mentioned strongly relevant dimension (figures will not be presented here due to space restrictions). This again supports our hypothesis that target object features modulate the search strategy. Further, the relevance analysis shows that an important part of the target feature impact may be focused on the fingertips.

References

1. Kohonen, T.: Self-organizing maps. Springer (2001)
2. Krieger, K., Moringen, A., Haschke, R., Ritter, H.: Shape features of the search target modulate hand velocity, posture and pressure during haptic search in a 3d display. In: Lecture Notes in Computer Science. Springer (2016), both authors contributed equally to this equally
3. Moringen, A., Haschke, R., Ritter, H.: Search procedures during haptic search in an unstructured 3d display. In: IEEE Haptics Symposium (2016)
4. Moringen, A., Krieger, K., Haschke, R., Ritter, H.: Haptic search for complex 3d shapes subject to geometric transformations or partial occlusion. In: IEEE World Haptics (2017)
5. Moringen, A., Aswolinskij, W., Buescher, G., Walck, G., Haschke, R., Ritter, H.: Modeling target-distractor discrimination for haptic search in a 3d environment (2018), (submitted to BioROB 2018), available under the following link http://www.techfak.uni-bielefeld.de/persons/abarch/biorob.pdf
6. Nowainski, J.: Efficient Target Identification during Haptic Search in a Three-Dimensional Environment. Bachelor thesis, Bielefeld University (2017), https://github.com/jnowainski/Bachelorarbeit/blob/master/tex/thesis/main.pdf
7. Schneider, P., Biehl, M., Hammer, B.: Adaptive relevance matrices in learning vector quantization (2009), https://pub.uni-bielefeld.de/publication/1994004