Space Perception by Visuokinesthetic Prediction

Wolfram Schenck and Ralf Möller

Computer Engineering Group, Faculty of Technology, Bielefeld University,
Bielefeld, Germany
wschenck@ti.uni-bielefeld.de

1 Introduction

It is a common assumption in embodied cognitive science that perception and cognition rely on the motor system [1–3]. In the "perception by anticipation" (PbA) approach [2], it is hypothesized that the visual perception of space and shape is based on an the internal simulation of a multitude of movement sequences. The predicted sensory outcomes of these sequences are thought to enable the perceptual process. It is supposed that the prediction in each single simulation step is performed by forward models (FMs) which anticipate the sensory consequences of motor commands.

Based on the PbA approach, we propose a robot model of space perception in a restricted domain in which a robot arm pushes a small block on a table surface (see Fig. 1, left). The model has two main components: first, a visuokinesthetic FM which predicts the visual image of the gripper tool and the kinesthetic state of the robot arm after a small movement step, and second an abstract recurrent network which associates the visual image of the gripper tool and the visual image of the block during pushing when they touch each other.

In our model, space perception means to perceive the location of the block on the table surface by generating a movement sequence which would move the gripper of the robot arm from its current position to a position where it would touch the block (as during pushing). This movement sequence is not executed, but just internally simulated. Thus, space perception is not linked to a metric coordinate system, but arises whenever the system knows how to *move* to the target object (here: the block).

The correct movement sequence is generated by an optimization process in which many sequences are tested in parallel. Each sequence has a final visual outcome (the visual image of the gripper tool after the last movement step as predicted by the FM). This outcome is overlayed with the current real image of the block, creating a composed visual state. Most of these states will show the gripper tool and the block at very different locations on the table surface; these visual states are irrelevant for space perception. Only if the (imagined) gripper tool and the block are as close to each other as during pushing, the corresponding movement sequence indicates the position of the block. To distinguish between irrelevant and relevant movement sequences, the abstract recurrent network (the visual memory) is used as novelty detector (novel overlayed visual state → irrel-

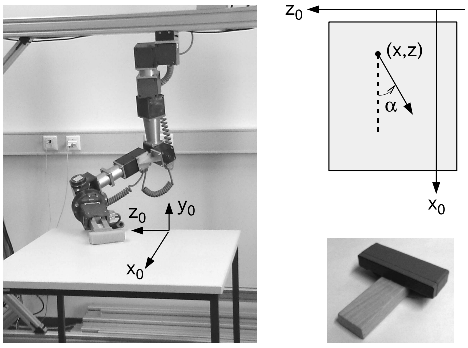

Fig. 1. Left: The robot arm in a pushing posture with the block in front of the gripper. Upper right: Base coordinate system on the table surface (see also left picture). The working area for pushing movements is shown in gray color. Lower right: Tool held by the gripper during pushing (adapted from [5]).

evant, non-novel → relevant). As optimization method, "differential evolution" (DE) [4] is applied, enforcing the minimization of the novelty.

2 Setup and Method

The used robot arm setup is shown in Fig. 1 (left). With the help of a special gripper tool (Fig. 1, lower right), the robot arm pushes a small wooden block on the table surface. This is the underlying motor capability which is not learnt but prewired. A camera records the image of the white table surface from above.

The visuokinesthetic FM is a multi-layer perceptron. As input, it receives the current kinesthetic state of the robot arm ($\mathbf{s}_{\text{KIN}}^{(t)}$), the current visual state related to the gripper tool ($\mathbf{s}_{\text{VG}}^{(t)}$), and a motor command for a small relative translational or rotational gripper movement ($\mathbf{m}_t$). As output, it estimates $\widehat{\mathbf{s}}_{\text{KIN}}^{(t+1)}$ and $\widehat{\mathbf{s}}_{\text{VG}}^{(t+1)}$ of the next time step. Both $\mathbf{s}_{\text{KIN}}$ and $\mathbf{m}$ are defined in the table base coordinate system (Fig. 1, upper right) in terms of the coordinates x and z and the orientation α. The training data for the FM comprises 37500 learning examples which were generated by systematically moving the gripper of the robot arm along different trajectories through the working area (while pushing the block) [5].

The visual memory for novelty detection is based on the NGPCA architecture [6]. It is adapted to training patterns which contain the vectors $\mathbf{s}_{\text{VG}}^{(t)}$ (visual state related to the gripper tool) and $\mathbf{s}_{\text{VB}}^{(t)}$ (visual state related to the block). This training data was also acquired during the abovementioned systematic pushing movements of the robot arm, thus the combined visual state $\mathbf{s}_{\text{VIS}}^{(t)} = \left(\mathbf{s}_{\text{VG}}^{(t)}, \mathbf{s}_{\text{VB}}^{(t)}\right)$ represents scenes in which the gripper tool always touches the block. Both $\mathbf{s}_{\text{VG}}$ and $\mathbf{s}_{\text{VB}}$ are generated from the camera image by color detection (green for the gripper tool and red for the block) and subsampling of the resulting image to

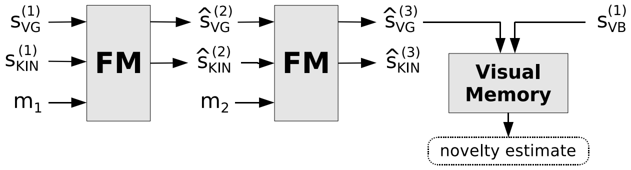

Fig. 2. The iterative application of the visuokinesthetic FM, depicted exemplary as chain of two FMs (with $t_0 = 1$ and $T = 3$). The initial sensory state is used as input to the chain, the final output $\widehat{\mathbf{s}}_{\mathrm{VG}}^{(3)}$ is combined with $\mathbf{s}_{\mathrm{VB}}^{(1)}$ as input for the visual memory. The estimated novelty indicates how strongly $\left(\widehat{\mathbf{s}}_{\mathrm{VG}}^{(3)}, \mathbf{s}_{\mathrm{VB}}^{(1)}\right)$ differs from visual states showing the gripper tool as it touches the block.

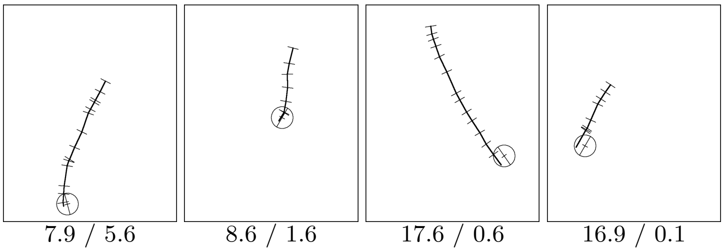

Fig. 3. Simulated trajectories for 4 different gripper and block positions/orientations (for details see text). The figures underneath each trajectory indicate the final position error (left; in mm) and the final orientation error (right; in degrees).

3 × 3 pixels, and by encoding the orientation of the gripper tool/block by a compass filter histogram with four values.

After learning, the FM and the visual memory are used for space perception as described in the introduction. Figure 2 illustrates the iterative prediction (from the current time step t_0 to the final time step T) and the application of the visual memory as novelty detector. The optimization method DE is used to find a movement sequence for which the novelty of the vector $\left(\widehat{\mathbf{s}}_{\mathrm{VG}}^{(T)}, \mathbf{s}_{\mathrm{VB}}^{(t_0)}\right)$ (the "overlay" of the predicted visual state of the gripper tool and of the real visual state of the block) is as small as possible. During optimization, 4050 movement sequences with a length between 7 and 15 steps are generated and tested.

3 Results and Discussion

Results To test the performance of our computational approach, we generated 100 perceptual tasks with random positions and orientations of the gripper and of the block (applying certain constraints to ensure that a simulated movement is geometrically possible). Each perceptual task was solved by the optimization process, finally yielding a sequence of motor commands. From this sequence, we computed the (hypothetical) gripper position and orientation after the last movement step. Ideally, this should equal the position and orientation of the gripper if it had pushed the block to its present location. On average, the position difference amounted to 15.1 mm (SD: 9.1), and the orientation difference to 3.7 degrees (SD: 4.2). For a workspace size on the table surface of 400 mm × 320 mm, this is a good performance. Figure 3 illustrates four perceptual tasks, showing the best simulated trajectory from the location of the gripper to the location of the block (indicated in each panel by a circle with a radius corresponding to 20 mm). Single movement steps are separated by small ticks.

Discussion In our approach, visual space perception is linked to the localization of objects by identifying a sequence of motor commands which would move the end effector from its current position to a position where it touches the object. The experimental results show for the tested task domain that this approach is successful in generating movements sequences with sufficient precision. It has not been experimentally verified yet, but it is not expected that human subjects show a better performance in a similar perceptual task.

The main components of the model, a visuokinesthetic FM and a memory for visual states, are in principle biologically plausible. The same holds for the visual overlay hypothesis since studies on mental imagery show that visual mental images have clear neural correlates in the visual cortex [7].

The model can be extended to space perception of objects which are not directly reachable by incorporating movements of the whole body. Moreover, if the end effector is not visible in the beginning, the simulated movement sequence could rely on a visuokinesthetic prediction which is only driven by kinesthetic inputs until the prediction provides a valid visual state for the end effector. In this way, the proposed model might be extended to a more general approach.

References

1. Gross, H.M., Heinze, A., Seiler, T., Stephan, V.: Generative character of perception: A neural architecture for sensorimotor anticipation. Neural Networks **12**(7-8) (1999) 1101–1129
2. Möller, R.: Perception through anticipation — a behavior-based approach to visual perception. In Riegler, A., Peschl, M., von Stein, A., eds.: Understanding Representation in the Cognitive Sciences. Plenum Academic / Kluwer Publishers, New York (1999) 169–176
3. Ziemke, T., Jirenhed, D.A., Hesslow, G.: Internal simulation of perception: A minimal neuro-robotic model. Neurocomputing **68** (2005) 85–104
4. Storn, R., Price, K.: Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4) (1997) 341–359
5. Schenck, W., Sinder, D., Möller, R.: Combining neural networks and optimization techniques for visuokinesthetic prediction and motor planning. In: ESANN 2008 proceedings. (to appear)
6. Möller, R., Hoffmann, H.: An extension of neural gas to local PCA. Neurocomputing **62** (2004) 305–326
7. Kosslyn, S.M.: Image and Brain. MIT Press, Cambridge, MA (1994)