Automated Learning of Coordinate Transformations

Xing Zhang and Michael W. Spratling
Division of Engineering
King's College London, Strand
UK, WC2R 2LS
xing.2.zhang@kcl.ac.uk michael.spratling@kcl.ac.uk

Eye-hand coordination is an important milestone of development both in infants and robots. It has been widely advocated that a robotic system can learn eye-hand coordination better through development rather than engineered designs (Brooks et al., 1999, Metta et al., 2000). The core of eye-hand coordination is automated learning of coordinate transformation, i.e. how to transform visually based coordinates (sensory input) to body based coordinates (motor information). The transformation is usually non-linear, which adds up the difficulty and needs prior knowledge of surrounding environment, which often means the transformational learning is biased (Metta et al., 1999, Kim et al., 2003, Cambron and Peters, 2001).

A novel model has been devised to learn automated coordinate transformations. As shown in Fig.1, the model is composed of two levels of neural networks. The lower level can transform retinocentric coordinates into head-centric coordinates, and the higher level can transform head-centric coordinates into body-centric coordinates. Each level of the network includes two layers of learning: conjunctive learning on the bottom and disjunctive learning on the top. The experiments on transforming retinocentric coordinates to head-centric coordinates have been conducted in a simulated problem and the results have shown that the transformation was successfully learnt, and can be a solid foundation for further developmental learning.

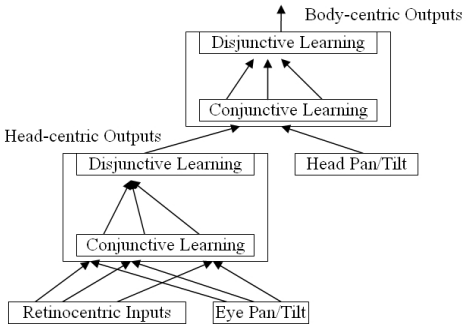

Figure 1: The structure of the model

1. The Model

The conjunctive layer can learn to represent conjunctions between different sensory inputs. In other words, when different sensory inputs are consistently coactive, this conjunction of the coactive inputs will be learned by the conjunctive layer.

In a visual system, when the camera varies it's pan/tilt position, a fixed object will appear correspondingly in different positions of the captured images. Therefore, for each pan/tilt position, there is a corresponding retinocentric position (in the image) for the fixed object. A conjunction can be formed by such pan/tilt position and retinocentric position. It is possible to vary the pan/tilt enough times to exhaust all such conjunctions. The conjunctive layer can learn all such conjunctions for the fixed object. If the fixed object is placed at various positions in the visual space, the conjunctive layer can exhaustively learn all such pan/tilt-retinocentric conjunctions of each position in the visual space.

The disjunctive layer attempts to learn the disjunctive relationships between sensory inputs. For a particular position in the visual space, it may have several pan/tilt-retinocentric conjunctions, but these conjunctions themselves are disjunctive to each other, i.e. no two conjunctions of a position will appear simultaneously in the sensory inputs. This is simply because one position will not occupy two or more positions in the physical world. The disjunctive layer will learn a set of outputs from the conjunctive layer which all represent the same position, and so form a head-centric representation.

Through such consecutive learning of conjunctions and disjunctions, the retinocentric coordinates can be transformed into head-centric coordinates. The same learning processes has been also carried out in the higher-level network to transform head-centric coordinates to body-centric coordinates.

2. Experiments

The model has been applied in a simulated experiment in learning coordinate transformation. The body-centric space is simulated by a 7×7 grid square,

with each grid representing a spacial point (Fig. 2). The head-centric space and the retinocentric space are simulated by a 5×5 square and a 3×3 square respectively. The smaller space can move around within the larger space and capture spacial points that have objects on them. Accordingly, the pan/tilt positions can be simulated by the position of the top-left corner of the smaller square.

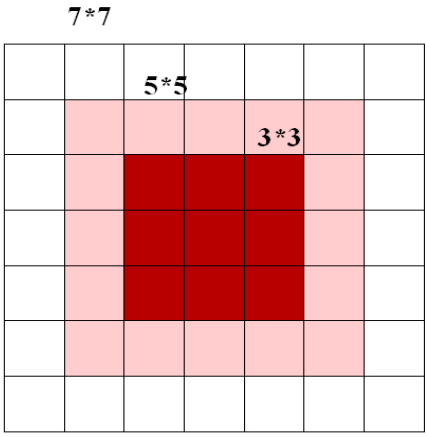

Figure 2: The simulated space

The model has been trained and tested on the artificial task and has successfully learnt all the conjunctions and disjunctions, which means body-centric representations have been formed from the retinocentric and pan/tilt inputs. The method of conjunctive and disjunctive learning is similar to methods that have previously been used to learn representations of objects with invariance to viewpoint (Spratling, 2005). Hence, the mechanisms that have been proposed to operate in the cortical ventral stream, are being used to learn coordinate transformation likely to develop in the dorsal stream.

Acknowledgements

We are grateful to EPSRC through grant EP/D062225/1.

References

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M. M. (1999). The cog project: Building a humanoid robot. *Lecture Notes in Computer Science*, 1562:52–87.

Cambron, M. and Peters, R. (2001). Determination of sensory motor coordination parameters for a robot via teleoperation. In *Proceedings of 2001 IEEE International Conference on Systems, Man, and Cybernetics*, volume 5, pages 3252–3257.

Kim, T. H., Kim, T. S., Dong, S. S., and Lee, C. H. (2003). Implementation of sensory motor coordination for robotic grasping. In *Proceedings of 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, volume 1, pages 181–185.

Metta, G., Panerai, F., Manzotti, R., and Sandini, G. (2000). Babybot: an artificial developing robotic agent. In *From Animals to Animats: the Sixth Int. Conf. on the Simulation of Adaptive Behavior (SAB2000)*, Paris, France.

Metta, G., Sandini, G., and Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12(10):1413–1427.

Spratling, M. W. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):753–761.