Wreath Product 3D Analysis and Movement Affordances: A Neurorobotic Approach

T.C. Henderson, T. Welker, and D. Sacharny
School of Computing
University of Utah
Salt Lake City, UT, USA
Email: tch@cs.utah.edu

***Abstract*— Symmetry-based wreath products have been proposed as a cognitive representation combining actuation and perception as derived from sensorimotor streams. Previous work has described how 2D wreath product shape analysis may be expressed in an appropriate neural network formulation, and here we extend this to include 3D surface analysis. That is, symmetry group operators with sensorimotor semantics are exploited in a neural network computational framework to provide a behaviorally relevant interpretation of a 3D scene. In addition, we describe how this approach provides a high-level abstraction which permits the recognition of similar geometric categories. Finally, we argue that this analysis can be implemented in a straightforward way as a neural network computation.**

I. Introduction

Gibson [4], [5] introduced the notion of *affordance* as something a physical feature offers or provides a perceiver in terms of possible action. The example he gives is that of a flat, horizontal, and rigid surface that affords support, thus, allowing mobility. Hartson [6] later refined the concept of affordance into several types, but labels this one a *physical affordance* and characterizes it as a physical relationship between an actor and physical aspects of the world which allow possible actions. We pursue this idea here wherein we propose to extend our previous work on wreath products as a cognitive representation [10], [11] to 3D behaviors for autonomous systems. In particular, we show how the wreath product allows for a representation which combines actuation (control) information with perception (sensed) information so as to allow the autonomous agent to act on the world in a coherent and valid way. The novel contribution here is the use of a neural network framework to exploit the wreath product representation so as to discover traversable parts of the environment.

Thus, the overall goal is to extract appropriate information from sensory data so as to allow an autonomous agent to act on that information in a useful (ecologically speaking) way. In previous work, we showed how 2D shape analysis could be effectively and efficiently expressed as a set of recurrent neural net subsystems [7].

Other work in this vein includes that of Almassy and Sporns [1] who propose a method for the perception of invariance which leads to categorization of visual stimuli based on the autonomous movement. Translation invariant representations are developed from the sensorimotor data resulting from the combined motor actions and concomitant sensory input. Their position is that biological systems require exploration and action in order to extract useful invariants from the environment, and furthermore, that success has a critical dependence on the physical structure of the organism.

Poggio and Anselmi [18] have explored invariant representations and some of their group theoretic aspects in the visual cortex and deep networks. They describe their results as follows:

> ... we develop a mathematical framework describing learning of invariant representations in the ventral stream. Our theory, called i-theory, applies to a broad class of hierarchical networks that pool over transformations. In particular, it applies to deep convolutional learning networks where the transformations are just translations in $\Re^2$. The networks associated with the theory do not have nonlinearities other than for pooling. In particular, i-theory applies to networks with pooling nonlinearities (to compute a histogram or associated statistics) such as sigmoidal threshold units or linear rectifiers ...

It has also been argued that the way from neural processing to behavior is not found by first deciding what to do and then figuring out how to do it [2], or even building a world representation separately from making action decisions. It may be the case that sensory data is processed in parallel, and that multiple actions are possible, and that these compete until a single, final action is selected. They propose that: "the dorsal visual system specifies actions which compete against each other within the fronto-parietal cortex, while a variety of biasing influences are provided by prefrontal regions and basal ganglia."

Most shape representations consider only sensed features of the object, e.g., geometry, color, measures of geometry like curvature or area, or relationships between such features. A generative description is proposed here which includes both actuation and sensory signals. This allows the shape to be synthesized by executing the motor controls specific to the shape. Leyton [15] proposed such a model for concept formation and uses the wreath product group (WP) [3] to encode the shape information. A wreath product captures the symmetries on a set of points and does this in terms of actions on subsets of points (these provide the connection to the agent's motor control system), and thus, these symmetries, represented as groups, must be extracted from the sensorimotor signals. This is a special form of affordance representation which exposes the interplay between perception and action in the world (see [4] for more on ecological psychology). To achieve an effective representation requires that: (1) symmetries be recognized in sensorimotor data, (2) that error and noise in the shape description be accounted for, and (3) that such an approach works within the agent's general cognitive framework. Wreath products are used to represent agent beliefs (restricted to geometry here). This approach has already been shown to be effective in the analysis of engineering drawings as well [10].

Earlier work on this topic includes the examination of innate theories as the foundation for cognition in robots [8], [9], as well as a method to characterize the representation of uncertainty using Bayesian methods [12], [13]. Those works provide concrete mechanisms to annotate the more abstract wreath products with well-defined coordinate frames to express the actuation required to generate the shape. Concepts then describe a specific thing (through the annotated information) or the general class of the shape (through the symbolic group structure of the WP).

More technically, a WP represents a shape as the semi-direct product of two groups – a fiber group which is acted upon by a control group. For example, the corner points of a square may be abstractly represented as $\{e\} \wr \mathcal{Z}_4$ or $\{e\} \times \{e\} \times \{e\} \times \{e\} \ltimes \mathcal{Z}_4$, where $\wr$ is the wreath product and $\ltimes$ is the semi-direct product symbol. Consider the second representation; four copies of a corner point (i.e., $\{e\} \times \{e\} \times \{e\} \times \{e\}$) are permuted by the cyclic group. Given a specific set of four points, say $S = \{[1,1]^T, [-1,1]^T, [-1,-1]^T, [1,-1]^T\}$, then the annotation includes S and the vector about which the points are rotated, e.g., $[0,0,1]^T$. Note that it may also be necessary to specify that $\mathcal{Z}_4$ has the same semantics as the group consisting of four rotations (i.e., $\{Rot_0, Rot_{\frac{\pi}{2}}, Rot_\pi, Rot_{\frac{3\pi}{2}}\}$) since there are several interpretations of the $\mathcal{Z}_4$ group. The annotation information allows the generation of the shape, and, in fact, this can be done in any new desired frame by the agent since annotations are relative to the agent's home coordinate frame. In terms of the agent's specific sensor and actuation mechanisms, it is supposed that there is a map from the 3D actuation to the specific agent motor systems; presumably, these maps would be learned as part of the embodied agent's developmental process. Thus, a practical, yet robust and powerful wreath product based cognitive system is provided as well as a demonstration of the basic affordance capabilities for interacting with the environment.

II. Motion Affordance

As described in the introduction, flat, horizontal, rigid surfaces afford motion to a mobile agent. This requires acquiring visual sensor data, recovering range (depth) information for that data, associating actuation with the perceived range data, and determining what motion is allowed and useful. Biological vision systems have been shown capable of producing stereo information, e.g., as disparity maps, etc., and we begin with the assumption that a range map (image) is available, e.g., as a set of values in a neural array.

Given a range image, $f(x,y)$, the surface $(x,y,f(x,y))$ is a Monge patch. From this we know that the surface normal, $\overline{n}$, can be found as follows:

$$\frac{\partial f(x,y)}{\partial x} = \frac{f(x+1,y)-f(x,y)}{\Delta x} = f_x(x,y)$$

$$\frac{\partial f(x,y)}{\partial y} = \frac{f(x,y+1)-f(x,y)}{\Delta x} = f_y(x,y)$$

Now consider the cross product of v_x and v_y, where:

$$\overline{v}_x = [x+1, y, f(x+1,y)] - [x,y,f(x,y)] = [1,0,f_x(x,y)]$$

$$\overline{v}_y = [x, y+1, f(x+1,y)] - [x,y,f(x,y)] = [0,1,f_y(x,y)]$$

$$\begin{aligned} \overline{n} &= \overline{v}_x \times \overline{v}_y \\ &= \begin{vmatrix} 0 & f_x(x,y) \\ 1 & f_y(x,y) \end{vmatrix} \bar{\mathbf{i}} - \begin{vmatrix} 1 & f_x(x,y) \\ 0 & f_y(x,y) \end{vmatrix} \bar{\mathbf{j}} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \overline{\mathbf{k}} \\ &= -f_x(x,y)\bar{\mathbf{i}} - f_y(x,y)\bar{\mathbf{j}} + 1 \cdot \overline{\mathbf{k}} \end{aligned}$$

This means that any two neighboring points in the same plane will have the same normal vector. Thus, in order to detect flat, horizontal, rigid surfaces, and agent must be able to ascertain that these three properties hold:

- *flat*: a flat surface is a planar surface and can be segmented by clustering the set of neighboring points in the scene that have the same surface normal.
- *horizontal*: in previous work on physical agents using wreath products for the cognitive representation of shape, we posited the knowledge of a gravity vector which indicates to the agent which way is down. This is used here to allow the recognition of a horizontal surface as having a normal vector whose dot product is -1 with the gravity vector.
- *rigid*: The determination of whether a surface is rigid or not is somewhat beyond the scope of our method, so we assume that if the agent is on a rigid surface at the current instant, then surfaces of similar appearance will be assumed rigid as well.

We do not address the issue of rigidity of the surface here, but seek flat, horizontal surfaces that afford motion (i.e., have a normal opposite to gravity). Note that for the present work, the chosen path is an element of the possible motions across a planar surface (as represented by a wreath product). Figure 1 shows the sequence of operations proposed.

This process can be implemented in neural network form as described in Figures 2 and 3. First, a Gaussian filter is convolved with the range image. Next, dx and dy filters are convolved. Convolution is easily produced in a neural network by simply placing the appropriate weights on the inter-layer connections. Given the gradient components, dx and dy, copies are made of each (simply map each neuron in the dx or dy layer to the corresponding neuron in the copy.

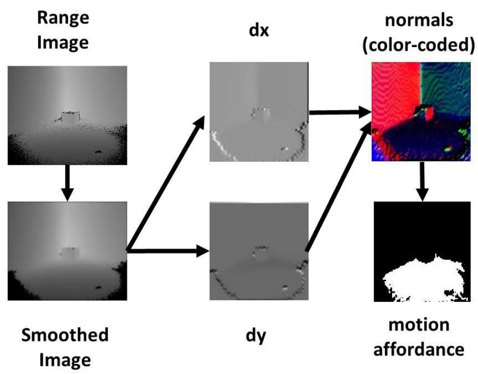

Fig. 1. Method to Determine Which Scene Elements Afford Motion. A range image is first smoothed, then the gradient is found; after this the unit surface normals are determined, and finally the acceptable motion area is extracted.

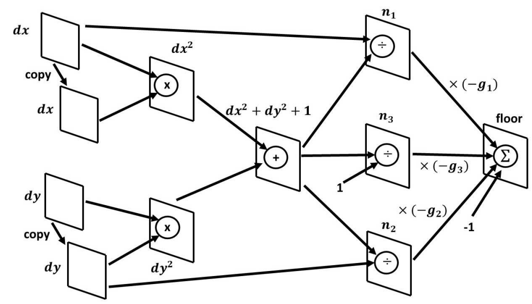

Fig. 3. Steps from Gradient to Motion Affordance Space.

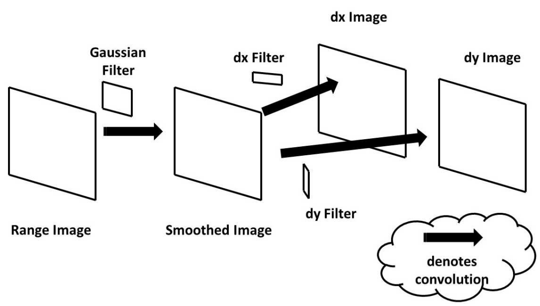

Fig. 2. Convolution Steps from Range Image to Gradient.

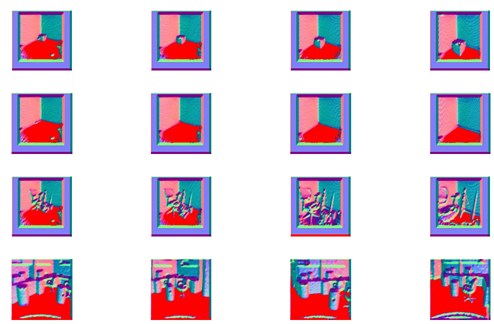

Fig. 4. Motion Afforded Areas in a Variety of Scenes. Row 1 shows the corner of a room with a box in the corner; from left to right, the Kinect sensor was moved closer; for all of these, the direction selected is straight ahead. Row 2 shows the room with no box and provides similar results as Row 1. Row 3 has some furniture in near the corner of the room; as can be seen, in the third column, there is not enough open floor to afford motion, whereas in the fourth column, some amount of floor becomes visible. Row 4 is an office scene, and in this case, the last two columns result in a decision to move to the left.

Next, the copies are multiplied pointwise with the originals (Siu et al. [19] describe how multiplication and division can be accomplished with a 4-layer neural network). Next the normal vector (i.e., [dx,dy,1]) length squared is determined by adding dx^2 with dy^2 and 1; this value is then used to produce a unit length normal vector ($[n_1, n_2, n_3]$), and finally, the dot product of the unit normal vector with the negative of the gravity vector is computed and compared to the value 1. That is, if the value in the *floor* array is zero, then the location is flat and horizontal and affords motion. A post-processing step (in terms of competing motion directions (this is in line with Cisek's proposal [2]) is performed to find the direction of the maximum number of such pixels (discretized into left, straight and right).

This process has been implemented (in Matlab), and the results on several types of scene are shown in Figure 4. Note that the colors in these images encode the normal vector directions (based on the neural network computation described above, except for *red* which indicates what parts of the scene afford motion).

A. Simulation Experiments

The motion affordance process has been tested in a 3D simulation environment. Figure 5 shows some box-like structures sitting on the x-y plane, and the patch of red dots provides the location of the range sensor array (note it is tilted down 45 degrees). Figure 6 shows a bit more complicated layout.

The motion affordance test procedure is as follows:

while $t < t_{max}$
    range $\leftarrow$ acquire range image
    normals $\leftarrow$ compute normals
    affordances $\leftarrow$ compare normals to gravity vector
    left $\leftarrow$ affordance strength left
    straight $\leftarrow$ affordance strength straight
    right $\leftarrow$ affordance strength right
    move $\leftarrow$ choose max strength direction
    increment t
end

Figures 7– 8, respectively, show the results of using the motion affordance process on the two layouts.

III. 3D Spatial Analysis

Wreath products also allow the representation of space as more abstract entities. For example, the symmetries of the

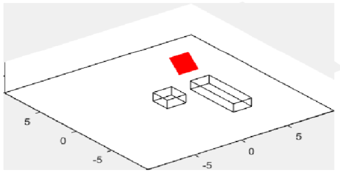

Fig. 5. Simulated Environment Layout 1.

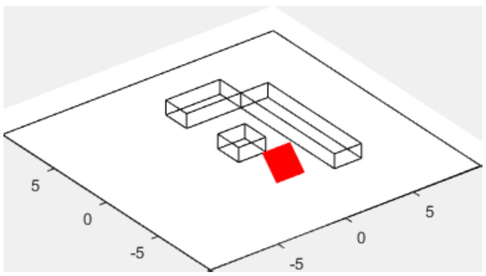

Fig. 6. Simulated Environment Layout 2.

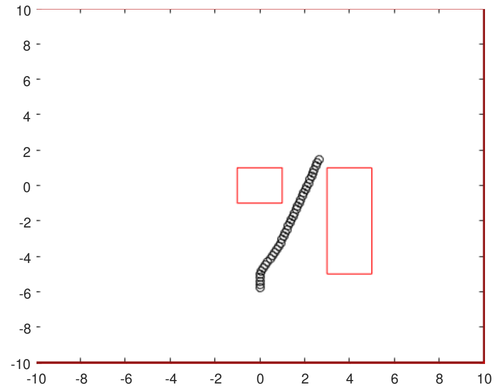

Fig. 7. Results of Motion Affordance Process in Layout 1.

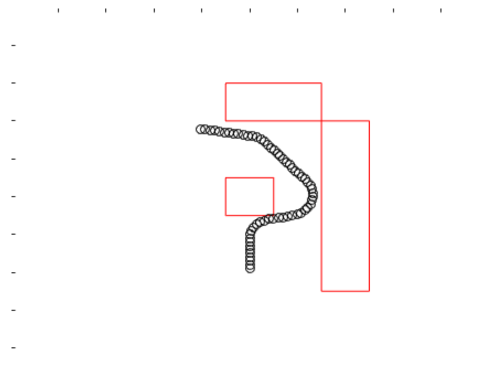

Fig. 8. Results of Motion Affordance Process in Layout 2.

cube may expressed as shown in Figure 9. Wreath products provide not just a description of the symmetries of the cube, but also an explicit plan for the production of the shape. In the case of the cube described in Figure 9, the lowest level group is a point, given as the identity group $\{e\}$ (note that in an implementation there is a corresponding annotation providing the coordinate frame and location of the point in that frame); this point is acted on (see [10], [15] for details on the wreath product and its application to the analysis of 3D data) by the translation group, $\Re$, to produce the side of a square; the side of the square may be transformed to any other side of the square by the cyclic group of order 4 ($\mathcal{Z}_4$), which in this case is the set of rotations by 0, 90, 180, and 270 degrees (again – about an appropriate axis); each square is related by a reflection symmetry to the square on the opposite side of the cube (Ref), and finally, there is a cyclic group order 3 symmetry ($\mathcal{Z}_3$) among the pairs of sides by the set of rotations (about a diagonal axis of the cube by rotations of 0, 120 and 240 degrees.

We now show how this wreath product may be detected in 3D range data in terms of a neural network that acts on a lower dimensional representation of the data. Given a range image (e.g., the first image on row 2 of Figure 4), the normal discontinuity image can be found by marking where the normal vector distance differs more than a specified threshold (see Figure 10 which shows the depth image of a room corner, and Figure 11 which shows the normal discontinuities for the empty room depth image). In previous work [7] we demonstrated a neural architecture to perform shape analysis by first producing the Frieze Expansion Pattern (FEP) [14], a log-polar form of the image, and then finding translation symmetries in the FEP. In the example here, the FEP is shown in Figure 12. A 1D translation (along the x-axis of the FEP) detects the $\mathcal{Z}_3$ symmetry in the data. This in turn guides the discovery of the visible sides of the cube. These matched groups in the wreath product template lead to the prediction of the remaining symmetries (i.e., the other sides of the cube). In addition, the wreath product tree easily permits the association of an associated Bayesian network to capture the uncertainties in the overall structure given the uncertainties in the data (e.g., range data noise).

Note that if 3D point data is available in a neural representation (e.g., as argued by Marr [17] in the 3D Sketch model), then wreath product symmetry analysis provides even more powerful methods for sensorimotor data based object representation. For example, give sample points on the interior surfaces of a cube (e.g., walls, ceilings, floor of a room), it is possible to perform a 3D Frieze Expansion Pattern by varying over two angles. This results in a reorganization of the cube data shown in Figure 13 which as can be seen, clearly brings out a set of 2D symmetries which characterize the shape of the cube. A Z_4 symmetry can be found (considering this depth image as a gray level image),

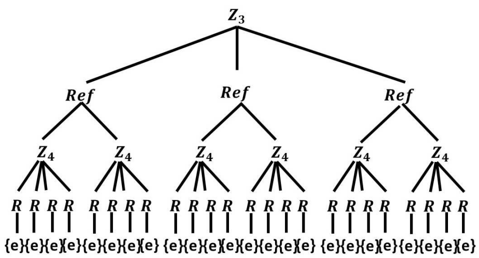

Fig. 9. One wreath product representation of a cube. The top
symmetry is a cyclic group of order 3 (rotations about a diagonal a
this moves pairs of reflected (Ref) faces (the $\mathcal{Z}_4$ symmetries) onto
pairs.

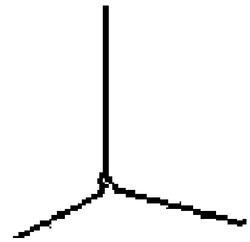

Fig. 11. Normal Discontinuity Image.

Fig. 10. Kinect Depth Image of a Corner of an Empty Room.

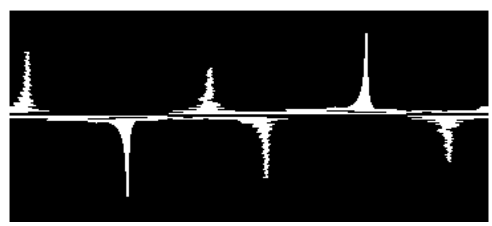

Fig. 12. The Log-Polar Image of the Normal Discontinuity Image.

and the translation axis for that is shown in the figure. Again, this is in line with our idea of reducing the dimensionality of the analysis space.

IV. Conclusions and Future Work

A novel approach to 3D surface motion affordance analysis is proposed based on the group theoretic wreath product. We have provided a neural network framework for its computation, and shown experimental results in a 3D simulation environment. Moreover, this approach fits well with our previously proposed wreath product based cognitive architecture for robot agents.

There are a number of directions for future work:

- Although the experiments performed here are based on computations analogous to neural networks, we intend to develop the motion affordance process in terms of a neuromorphic architecture. This will allow more insight into the computational efficiency of the method in a pure neural inspired framework.
- Once a neural computational element is available, we propose to explore its effectiveness in the Neurorobotics Platform (NRP) [16] being developed by the Neurorobotics Subproject of the European Human Brain Project. This platform provides (1) Robot Designer, (2) Environment Designer, (3) Experiment Designer, and (4) Virtual Coach components to help develop and test processes like ours. The NRP makes it easy to couple neural networks to robots.
- We have proposed a general cognitive architecture based on wreath products, and we intend to explore its use in a mobile embodied system in which motion affordance will play a key role.

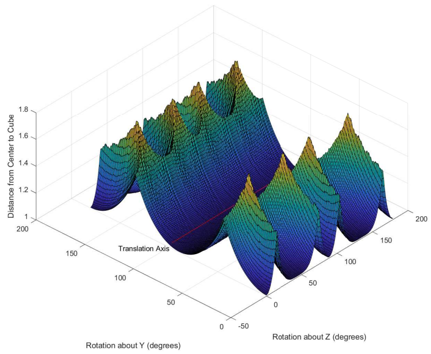

Fig. 13. The Frieze Expansion Pattern of a 3D Cube using a 2 Angle Sweep.

REFERENCES

[1] N. Almassy and O. Sporn. Perceptual Invariance and Categorization in an Embodied Model of the Visual System. In B. Webb and T.R. Consi, editors, *Biorobotics*, Menlo Park, CA, 2001.

[2] P. Cisek. Cortical Mechanisms of Action Selection: the Affordance Competition Hypothesis. *Philosophical Transactions B*, 362(1485):1585–1599, September 2007.

[3] D.S. Dummit and R.M. Foote. *Abstract Algebra*. Wiley & Sons, Hoboken, NJ, 2004.

[4] J.J. Gibson. The Theory of Affordances. In R. Shaw and J. Bransford, editors, *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, pages 67–82, Hillsdale, NJ, 1977. Lawrence Erlbaum.

[5] J.J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin Co, Boston, MA, 1979.

[6] Rex Hartson. Cognitive, Physical, Sensory and Functional Affordances in Interaction Design. *Behaviour & Information Technology*, 22(5):315–338, September 2003.

[7] T.C. Henderson and T. Beall. A Sensorimotor Approach to Concept Formation using Neural Networks. In *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, Baden-Baden, Germany, September 2016.

[8] T.C. Henderson, A. Joshi, and E. Grant. From Sensorimotor Data to Concepts: The Role of Symmetry. Technical Report UUCS-12-012, The University of Utah, Salt Lake City, UT, October 2012.

[9] T.C. Henderson, H. Peng, C. Sikorski, N. Deshpande, and E. Grant. Symmetry: A Basis for Sensorimotor Reconstruction. Technical Report UUCS-11-011, The University of Utah, May 2011.

[10] A. Joshi. *Wreath Product Cognitive Architecture*. PhD thesis, University of Utah, Salt Lake City, UT, 2016.

[11] A. Joshi and T.C. Henderson. Wreath Product Cognitive Architecture (WPCA). In *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, Baden-Baden, Germany, September 2016.

[12] A. Joshi, T.C. Henderson, and W. Wang. Robot Cognition using Bayesian Symmetry Networks. In *Proceedings of the International Conference on Agents and Artificial Intelligence*, Angers, France, 2014.

[13] A. Joshi and Thomas C. Henderson. Generative cognitive representation for embodied agents. In *Multisensor Fusion and Information Integration (MFI)*, Beijing, China, 2014.

[14] S. Lee, R. Collins, and Y. Liu. Rotation Symmetry Group Detection via Frequency Analysis of Frieze-Expansions. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[15] M. Leyton. *A Generative Theory of Shape*. Springer, Berlin, 2001.

[16] A. Mahmoud. The Neurorobotics Platform: A Simulation Environment for Brain-inspired Robotics. In *The Bernstein Conference*, Goettingen, Germany, September 2017.

[17] D. Marr. *Vision*. W.H.Freeman, San Francisco, CA, 1982.

[18] T.A. Poggio and F. Anselmi. *Visual Cortex and Deep Networks*. MIT Press, Cambridge, MA, 2016.

[19] K.-Y. Siu, J. Bruck, T. Kailath, and T. Hofmeister. Depth Efficient Neural Networks for Division and Related Problems. *IEEE Transactions on Information Theory*, 39(3):946–956, May 1993.