Modeling Perspective-Taking by Correlating Visual and Proprioceptive Dynamics

Fabian Schrodt[1] (tobias-fabian.schrodt@uni-tuebingen.de)
Georg Layher[2] (georg.layher@uni-ulm.de)
Heiko Neumann[2] (heiko.neumann@uni-ulm.de)
Martin V. Butz[1] (martin.butz@uni-tuebingen.de)

[1] Department of Computer Science, Chair of Cognitive Modeling
University of Tübingen, Sand 14, Tuebingen, Baden-Württemberg, 72076 Germany

[2] Faculty of Engineering and Computer Sciences, Institute for Neural Information Processing
Ulm University, James-Franck-Ring, Ulm, Baden-Württemberg, 89081 Germany

Abstract

How do we manage to step into another person's shoes and eventually derive the intention behind observed behavior? We propose a connectionist neural network (NN) model that learns self-supervised a prerequisite of this social capability: it adapts its internal perspective in accordance to observed biological motion. The model first learns predictive correlations between proprioceptive motion and a corresponding visual motion perspective. When a novel view of a biological motion is presented, the model is able to transform this view to the closest perspective that was seen during training. In effect, the model realizes a translation-, scale-, and rotation-invariant recognition of biological motion. The NN is an extended adaptive resonance model that incorporates self-supervised error backpropagation and parameter bootstrapping by neural noise. It segments and correlates relative, visual and proprioceptive velocity kinematics, gradually refining the emerging representations from scratch. As a result, it is able to adjust its internal perspective to novel views of trained biological motion patterns. Thus, we show that it is possible to take the perspective of another person by correlating proprioceptive motion with relative, visual motion, and then allowing the adjustment of the visual frame of reference to other views of similar motion patterns.

Keywords: Proprioception; vision; associative learning; self-supervised learning; mental rotation; canonical views; point-light display; biological motion; recurrent neural networks.

Introduction

Do we build up and use our own sensorimotor knowledge to adapt our mental perspective for putting ourselves into another person's shoes? We present an artificial neural network (NN) model that is able to do just that: it enables a view-point independent perception of biological motion, correlating relative visual motion to corresponding proprioceptive motion, and later mentally transforming novel visual motion perspectives to previously learned, canonical perspectives. In several fields of research, the ability to mentally transform the own coordinate system to match an observed one is referred to as self-projection, view-point adaptation, or perspective-taking. To enable such a mental transformation, we focus here on the perception of biological motion.

The perception of biological motion is assumed to be related to the Superior Temporal Sulcus (STS) (Pavlova, 2012; Pyles et al., 2007), where temporal and parietal pathways of visual information processing converge. However, since the parietal path is also strongly involved in the perception of body-relative spaces (Holmes & Spence, 2004), it seems likely that parietal visuo-proprioceptive correlations are learned in early mental development. We investigate how such self-induced correlations can be learned and how this knowledge can be used to adapt the visual processing of observed actions. As Johansson pointed out, showing a moving set of light points representing the locations of a walking person's joints is sufficient to perceive the underlying biological motion (1973). Thus, it seems that this ability is at least partially based on the perception of the relative motion of bodily feature locations. Hence, we use the motion of relative feature locations as the visual input to our model. Seeing that the noise-tolerance of subjects identifying biological motion from point-light walkers decreases dramatically if the presentation is inverted top-down (Pavlova & Sokolov, 2003), it appears that canonical views of motions affect biological motion recognition similar to how canonical views of objects influence object recognition (Shepard & Metzler, 1971). Thus, we propose an NN model that learns canonical views of biological motion, later adjusting the internal frame of reference to deduce another person's perspective while monitoring her or his biological motion patterns.

The neural network we propose segments a continuous sensory stream in a common visual and proprioceptive space at nonlinearities in the coincident motion dynamics. Persistent spatiotemporal congruencies in this domain are memorized by Hebbian-inspired learning rules. The model is scale and translation invariant, because it processes directions of relative velocities in the considered sensory spaces. Assuming that joint angles can be considered a part of the proprioception that can also be perceived visually, a view-point invariant perception of observed biological motion is enabled by rotating visual feature locations in accordance to proprioceptive experience. The model realizes this by top-down error projections that minimize the divergence between perceived positional motion and the previously associated angular motion. We show that our model is able to progressively adapt its internal perspective on visual data in a self-supervised manner, effectively slipping into another person's shoes. In sum, we present a model that is able to derive another person's perspective by exploiting sensorimotor knowledge about the own body kinematics. Since it seems necessary to know another

person's perspective to some degree for deriving her or his current intentions, the cognitive capabilities we investigate may set the stage for the development of the mirror neuron system. Thus, the model offers one highly embodied, developmental path for bootstrapping the capability to imitate another person, to derive her or his intentions, and even to be empathetic.

A related NN modeling STS was developed previously (Layher et al., 2014) to fuse visual information from motion and form pathways. This model did not include proprioceptive dynamics and was not able to adjust the internal perspective. Another recent NN architecture modeled view-independence during object interactions (Fleischer et al., 2012), but it did neither *learn* correlations nor canonical views, and it did not investigate the adjustment of the internal perspective, either.

The remainder of the paper is structured as follows: First, we introduce our point-light simulation environment to describe example inputs to the model. Then, we introduce the neural architecture for matching dynamics, building canonical views, and progressive mental transformation. In this respect, we also explain how to flexibly bootstrap the network's weights on the basis of noise without prior knowledge about the input data. Next, we evaluate the model in three experimental setups. Finally, we summarize the results, sketch-out implications, and point to future research options.

Simulation Environment

To exemplify a setup of our model, we implemented a simple 2D, 2DOF arm-simulation (see Fig. 1). The arm executes a continuous forward and backward swing, somewhat similar to the arm of a walking person. The simulation provides the relative locations of all joints in retinal coordinates $(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)$, as well as the (proprioceptive) shoulder and elbow joint angles (α, β). Visual information can be transformed by an experimenter (that is, rotated, mirrored, etc.) before serving as input for the neural network model. The corresponding visual rotation angle of the entire arm is denoted ν in the following.

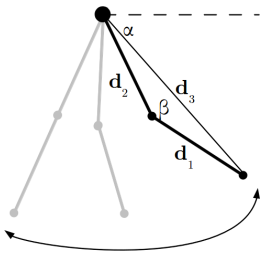

Figure 1: Point-light arm simulation. The big dot represents the shoulder location, followed by elbow and wrist.

Neural Network Model

The model consists of three successive stages illustrated in the overview given in Fig. 2. The first stage processes visual and proprioceptive information. That is, it visually receives relative locations of joints, and absolute joint angles by the proprioceptive system. Subsequently, stage I convolves these data separately into directional velocities. In this process, mental rotation is applied to the visual information. Note that the angular information is rotation invariant and may thus also be derived from vision without transformation.

Stage II performs a modulatory normalization of information and then pools the information from the visual and proprioceptive streams of information. Stage III implements a self-supervised adaptive resonance model. It uses instar-learning to segment the sensory stream given by stage II and to memorize permanent correlations, and outstar-learning to recall or predict the learned correlations. Compared to the classical unsupervised approach, this allows to derive a prediction error, which is backpropagated through the network.

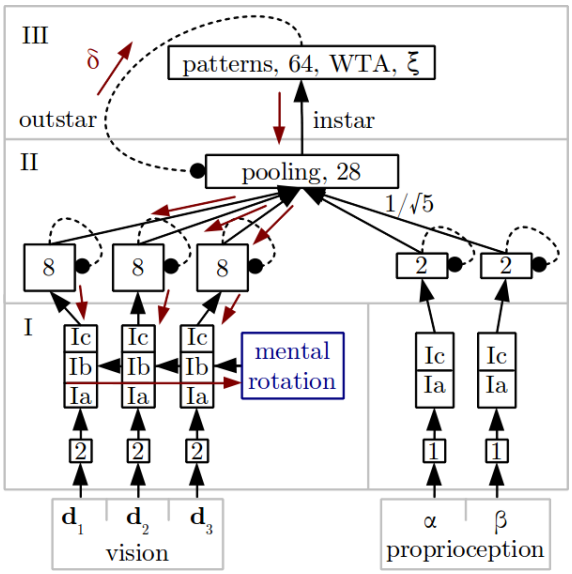

Figure 2: Overview of the three-stage neural modeling approach. Boxes numbered with n indicate layers consisting of n neurons. Black arrows describe weighted forward connections between layers, while bullet-heads indicate modulations. Dashed lines denote delay-connections. Red arrows denote the backward propagation of prediction errors.

Stage I - Feature Processing

The visual input path of the network is driven by relative (2D) coordinates of simulated arm joints. Analogously, the proprioceptive path of the model is driven by the (1D) angles between limbs. We momentarily assume that the coordinates of hand, elbow and wrist and their according angles can be identified and assigned to the respective input neurons reliably. We chose the angular information as the only proprioceptive input, and assume that this information can also be derived from vision upon action observation. Fig. 3 shows the feature processing for a single, two-dimensional visual limb relation in the visual path. The input is, for example,

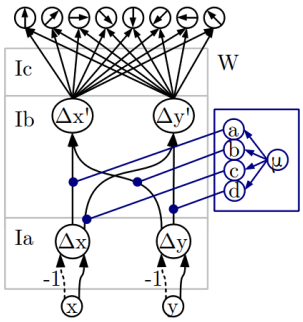

Figure 3: Processing path of a relative joint location

the hand-location minus the elbow-location $\mathbf{d}_1 = \begin{pmatrix} x & y \end{pmatrix}^T$ in retinal coordinates.

In interstage Ia, this information is transformed into a directional velocity by time-delayed subtraction. In this way, the model becomes translation-invariant. In interstage Ib, the information of directional velocity is transformed by means of a presynaptic, gain field-like modulation (Andersen et al., 1985), simulating a mental transformation. It is realized by a two dimensional matrix multiplication of the directional velocity $\begin{pmatrix} \Delta x & \Delta y \end{pmatrix}^T$ into a transformed directional velocity $\begin{pmatrix} \Delta x' & \Delta y' \end{pmatrix}^T$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} \Delta x' \\ \Delta y' \end{pmatrix}.$$

When the rotation applies, the neurons a,b,c,d implement the elements of the rotation matrix, which is driven by an adjustable rotation angle μ realized by a bias neuron. While the same mental rotation angle is applied to all visual information, no mental rotation is applied to the single-dimensional angular pathway. Interstage Ic implements directional convolution over time, converting the directional velocities into direction-responsive activities. The weighing matrix is set up in a combinatorial fashion, as every single dimension of the feature input may increase, not change, or decrease. W is furthermore normalized per direction. This procedure can easily be performed for features of every dimensionality D, resulting in $3^D - 1$ direction sensitive neurons.

In all, stage I provides a population of neurons for each sensory feature processed, which is either sensitive to directional velocities in the visual position of a feature (8 neurons) or sensitive to directional velocities of a proprioceptive angle (2 neurons).

Stage II - Normalization and Pooling

Stage II firstly accounts for a separate normalization of activity in the direction-sensitive populations, which is indicated by feedback connections in Fig. 2. In this way, absolute velocities are ignored and only the directions of changes in visual/proprioceptive information are taken into consideration, by which the model becomes scale-invariant. Normalization of a layer's activity-vector can be accomplished by axo-axonic modulation. The method we propose approximates a real-time normalization of a layer's output-vector to the Euclidean length 1. In our model, a common neuron indexed by j can formally be described by its input net_j, its activation function $f_j(\text{net}_j)$, its output o_j, some noise-term ξ_j (which we will address later) and the axonic modulatory factor a_j:

$$o_j = a_j \cdot f_j(\text{net}_j) \tag{1}$$

$$\text{net}_j = \xi_j + \sum_i w_{ij} \cdot o_i \ . \tag{2}$$

Normalization of the neural activity in a layer a is realized by modulating all neurons j of that layer by the output o_a of a single, layer-specific normalizing neuron ($a_j := o_a$)[1], with

$$o_a(t) = \frac{o_a(t-1)}{\sum_j \overline{o_j}(t)^2} \ , \tag{3}$$

where $\overline{o_j}$ denotes a delayed moving average of the output of neuron j:

$$\overline{o_j}(t) = (1-\lambda) \cdot \overline{o_j}(t-1) + \lambda \cdot o_j(t-1) \tag{4}$$

with decay parameter $\lambda \in (0,1]$.

After normalization, all direction-sensitive fields are pooled by one-to-one connections into a single, bigger pooling layer, which serves as input to stage III. The connections are weighted by $1/\sqrt{n}$, where n denotes the number of sensory information sources being processed (5 in our example). In this way, also the pooling layer input is normalized, which is important for the applied learning rules.

Stage III - Correlation Learning

Stage III realizes a segmentation of the normalized and pooled information from stage II (neurons indexed by i) by means of a number of pattern responsive neurons (indexed j, quasilinear in the range $[0,1]$). Each pattern neuron becomes the representative for a unique constellation of positional and angular directions of variability. For segmentation, we use instar learning (Grossberg, 1976a):

$$1/\eta \cdot \partial w_{ij}(t)/\partial t = \Delta w_{ij}(t) = o_j(t) \cdot (\text{net}_i(t) - w_{ij}(t)) \ , \tag{5}$$

with learning rate η. The rule implies that the weight vector $\mathbf{w}_j$ to each single pattern neuron j approaches the input vector of the preceding layer at a rate determined by the pattern's activity. To avoid "catastrophic forgetting" of patterns, we use winner-takes-all (WTA) competitive learning (Rumelhart & Zipser, 1985) in the sense that only the weights to the most active neuron in the pattern layer are adapted.

Grossberg's "sparse patterns theorem" (Grossberg, 1976b) states that learned patterns can in general only be guaranteed to be stable if the initial weight vectors underlie a certain distribution, which depends on the actual subspace of input vectors. Since the input space is initially typically unknown, we bootstrap the weight vectors from scratch ($w_{ij}(t_0) = 0$) by a

[1] No square root is necessary for the normalization to length 1.

neural noise mechanism. Initially, no sensory information is propagated to the pattern layer ($o_j(t_0) = 0$). Instead, we add normally distributed noise $\xi_j = \mathcal{N}(0, \sigma)$ to the input net_j of each neuron in the pattern layer (see Eq. 2), such that pattern neurons are driven by the sum of signal and noise. Thus, some random pattern neuron is initially the winner, and its weight vector adapts from 0 to a novel pattern.

To account for the issue that the weight vectors are not being normalized to a length of 1 in this way – which is an important property of instar-learning – we assume that the excitability of a pattern neuron decreases proportional to its overall synaptic strength:

$$f_j(\text{net}_j) = \text{net}_j \cdot \min(||\mathbf{w}_j||^{-1}, r) \ , \tag{6}$$

where r denotes the upper excitability boundary. By that, the weight vector to a pattern neuron is normalized if its length exceeds r. In contrast to other approaches, this procedure does not initialize the pattern neurons' instar weights with a random direction but changes their response randomly, and with it, their probability to win. During the development of a pattern, the winning probability is magnified by the angle between the presented pattern and the weight vector of a pattern neuron, while the relative influence of neural noise decreases. Thus, both the amount of neural noise – determined by σ – and the initial responsiveness r play a major role for the distribution of the network's pattern capacity: While σ influences the probability that a developed pattern is retrained, $\sigma \cdot r$ determines the probability that an undeveloped pattern wins over a developed one and is thus consulted to increase the spatial resolution of this episode in dynamics.

We furthermore use predictive outstar learning as an attentional gain control mechanism, by which stage III becomes an adaptive resonance (or self-stabilizing) model (Grossberg, 1976c). This is realized by feedback connections from the pattern layer to the pooling layer, which are trained by

$$1/\eta \cdot \partial w_{ji}(t)/\partial t = \Delta w_{ji}(t) = o_j(t-1) \cdot (\text{net}_i(t) - w_{ji}(t)) \ , \tag{7}$$

where neuron j is the winner of time step $t-1$. This means that the outgoing weight vector of the last most active pattern neuron also approaches the input of the pooling layer (which again activates the neuron itself) and thereby learns a prediction over a marginal time span.

The absolute outstar learning signal is also used on forward propagation from the winner of time step $t-1$ as axo-axonic modulatory gain in the pooling layer i:

$$a_i(t) = 1 - \left|\Delta w_{ji}(t)\right| \in [0,1] \ . \tag{8}$$

By this modulation, the last winner inhibits the pooling layer's output (via Eq. 1): the larger the error in and the larger the reliability of the prediction, the stronger is the resulting inhibition (cf. Eq. 7). In result, the pattern distinction is improved further.

The prediction error is in turn also being backpropagated top-down through the network to adapt the mental transformation in an error-minimizing manner (see Fig. 2). In our model, the perspective adaptation is thus driven by the visual kinematics expected for the proprioceptive dynamics and vice versa. The prediction error δ_j, which is backpropagated over the outstar weights to a pattern neuron, can be described by the weighted sum of negative prediction deviations in the pooling layer[2]:

$$\delta_j(t) = \sum_i w_{ji}(t) \cdot (w_{ji}(t) - \text{net}_i(t)) \ , \tag{9}$$

where j is again the winner neuron of time step $t-1$. This is equivalent to

$$\delta_j(t) = 1 - \sum_i \text{net}_i(t) \cdot w_{ji}(t) \ , \tag{10}$$

under the assumption that the outstar weights have completed training (hence have length 1). Thereby, the error of a pattern neuron is determined by the angle between the predicted and the actual pooled information. Assuming furthermore that the predictions have been learned correctly, the error-driven adaptation of the network's parameters maximizes the response of patterns that represent the momentary constellation of positional and angular dynamics best.

Experiments

In the following, we evaluate our NN model on psychological findings. The network architecture was parametrized with $\sigma = 0.002$, $r = 100$, a learning rate of $\eta = 0.04$ for instar/outstar and mental rotation learning, and with 64 neurons in the pattern recognition layer. We took the average of 100 independent runs for all experiments.

Baby-Mirror-Test

In a psychological experiment done by Rochat and Morgan (1995), two real-time videos with different spatial transformations were presented to infants, showing their own leg-movements. The infants paid significantly more attention to a mirrored view of their movements than to an untransformed presentation. This underlines the importance of movement directionality in self-perception.

In our experiment, we interpret the attention of an infant to be guided by surprise (see e.g. Itti & Baldi, 2006), or an error in the prediction of visual feedback, assuming that a normal self-perspective has been learned. To show that an exogenous visual transformation produces such an error, we trained our model on the specified arm simulation. The forward and backward swings of the arm were clearly distinguished after presenting the whole movement (forward and backward) 300 times. We then decorrelated the visual feedback from the associated proprioception by left-right inversion of the network's visual input. As can be seen in Fig. 4, the prediction error (RMS of the winner pattern neuron) rises and stays constant since we neither allow further learning nor

[2] All other backpropagation terms (including those for gain fields) and the weight adaptation rules for the model's internal perception angle μ follow from gradient descent.

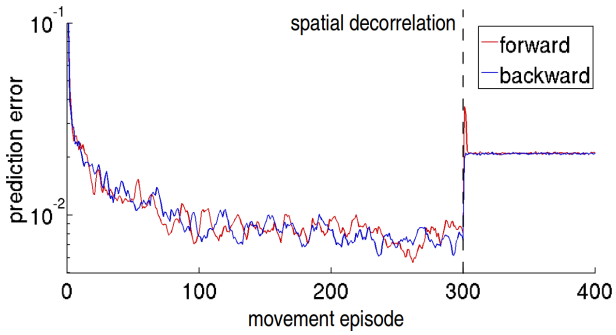

Figure 4: Baby-mirror-test. The prediction error decreases during training, and increases afterwards when the simulated vision is left-right inverted.

the adjustment of the model's internal perspective μ, effectively simulating the surprise of the babies when confronted with an inverted video display.

Canonical Views

View-based representations of goals and biological motion have been found in the macaque premotor cortex area F5 (Caggiano et al., 2011) as well as in the (posterior) STS, respectively, which are both considered to be part of – or contributing to – the mirror neuron system. Those view-dependent cells are assumed to play a part in the resulting view-independence of action recognition associated to further cells found both in F5 (Caggiano et al., 2011) and (anterior) STS (Jellema & Perrett, 2006). We show that our model is able to learn multiple view-dependent representations of biological motion, which we term 'canonical views'.

In this experiment, we trained the model on three rotated perspectives ($\nu \in \{0, 120, -120\}°$) of the arm movement and repeated that training 4 times. 50 full arm movements were presented in each perspective, resulting in 600 full arm swings altogether. As in the baby mirror test, we did not allow the adaptation of μ.

Fig. 5 shows that multiple canonical views could generally be learned by the same network without relearning patterns in between: The bar plot counts the number of pattern neurons

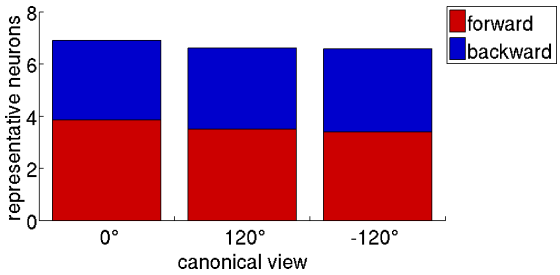

Figure 5: Separate representation of canonical views and motion directions. The three canonical views and two motion directions (color-coded) are learned in six disjunct groups of pattern neurons.

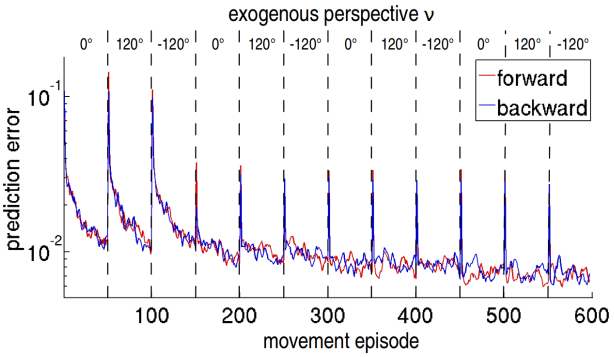

Figure 6: Learning multiple canonical views. The prediction error decreases for each canonical view trained repeatedly. The smaller peaks in the error when the same view is shown again indicate pattern recognition.

that were exclusively winning in all repetitions of a single perspective, stacked for the forward and backward swing. Both the forward and backward swings as well as the canonical views of the whole movement were represented by a comparable amount of patterns, while movements learned early were slightly favored in terms of the number of patterns.

Fig. 6 shows the trend of the prediction error over time. It can be seen that the prediction error decreased separately for each trained canonical view. Thus, several, independent canonical views can be learned and maintained by the NN model.

Mental Transformation

Motivated by the fact that humans appear to mentally rotate objects to their respective closest, known canonical view (Shepard & Metzler, 1971), in the final experiment we investigate if the model is able to transform biological motion to the closest canonical motion view. In order to show that the model is able to change its internal perspective using the expected directional correlations, we set the exogenous rotation ν of the visual feedback to a random value within $[-180, 180]°$ after learning three canonical views as in the last experiment. In doing so, we did not allow new patterns to arise, but allowed the model/mental rotation μ to adapt according to the prediction error backpropagated to the mental rotation module via the individual Ib stages in the visual pathway.

Fig. 7 shows the adaptation of the overall rotation $(\nu + \mu)$ over time for all tested trials. The model adapted its mental rotation angle μ progressively to the nearest (in terms of orientation difference) canonical view that was learned before without explicit knowledge about the simulation angle ν. Thus, the NN model is able to derive the perspective of another person (in this case a simple arm) by learning to associate visual motion of relative joint locations with the angular motion of the joints. This adaptation is driven by the error in self-generated predictions.

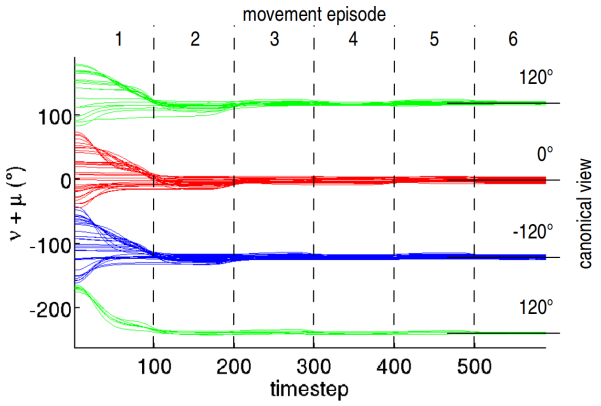

Figure 7: Perspective taking. When a novel view on the trained motion is shown, the perspective gradually converges to the nearest canonical view.

Summary, Conclusion & Future Work

The presented results have shown that our NN model is able to simulate (a) surprise when being presented with unexpected directionalities in visual kinematics; (b) the learning of multiple canonical views of biological motion and thus the generation of both, view-dependent and view-independent visual mirror neurons; (c) the prediction-error-driven adaptation of the visual perspective to derive the perspective of another person.

Despite the rather large degree of abstraction in our model, our experiments confirm that directionalities in a visuoproprioceptive space alone may suffice to put oneself into another person's shoes. While the brain may certainly use other clues as well, it appears that the perception of biological motion plays a crucial factor (Pavlova, 2012). In effect, our model offers an embodied pathway towards learning mirror neuron capabilities, imitating other people, deriving their intentions, and even showing empathy. Further model evaluations may even yield implications for understanding social dysfunctions, such as autism.

Despite the capabilities of the model, further investigations are necessary. Most importantly, here we assigned bodily features to neural inputs directly. However, when visual input is presented, the observed features still need to be properly mapped to the respective body parts and thus to the corresponding neural inputs. Also, additional information sources may be considered such as motor activity, the axis of gravity, information about the floor / the ground, acceleration, or further visual features about the observed body. Moreover, the capability of dealing with information missing or distorted on action observation needs to be further investigated. Finally, along with object-relative information, a model for self-supervised and view-independent imitation learning could be established.

Acknowledgments

GL and HN have been supported by the SFB Transregio 62 funded by the German Research Foundation (DFG).

References

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, *230*(4724), 456–458.

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., et al. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Current Biology*, *21*(2), 144–148.

Fleischer, F., Christensen, A., Caggiano, V., Thier, P., & Giese, M. A. (2012). Neural theory for the perception of causal actions. *Psychological research*, *76*(4), 476–493.

Grossberg, S. (1976a). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics*, *21*(3), 145–159.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. *Biological cybernetics*, *23*(3), 121–134.

Grossberg, S. (1976c). Adaptive pattern classification and universal recoding: II. feedback, expectation, olfaction, illusions. *Biological cybernetics*, *23*(4), 187–202.

Holmes, N. P., & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing*, *5*, 94-105.

Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. *Advances in neural information processing systems*, *18*, 547.

Jellema, T., & Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia*, *44*(9), 1535–1546.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, *14*(2), 201–211.

Layher, G., Giese, M. A., & Neumann, H. (2014). Learning representations of animated motion sequencesa neural model. *Topics in Cognitive Science*, *6*(1), 170–182.

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex*, *22*(5), 981-995.

Pavlova, M. A., & Sokolov, A. (2003). Prior knowledge about display inversion in biological motion perception. *Perception*, *32*(8), 937–946.

Pyles, J. A., Garcia, J. O., Hoffman, D. D., & Grossman, E. D. (2007). Visual perception and neural correlates of novel biological motion. *Vision Research*, *47*(21), 2786–2797.

Rochat, P., & Morgan, R. (1995). Spatial determinants in the perception of self-produced leg movements in 3- to 5-month-old infants. *Developmental Psychology*, *31*(4), 626-636.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive science*, *9*(1), 75–112.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.