Neural Model for the Influence of Shading on the Multistability of the Perception of Body Motion

Leonid Fedorov[1 2], Joris Vangeneugden[3] and Martin Giese[1 2]
[1]*Dept. of Cognitive Neurology, CIN, HIH, University Clinic Tuebingen, Tuebingen, Germany*
[2]*IMPRS for Cognitive and Systems Neuroscience, University of Tuebingen, Tuebingen, Germany*
[3]*School of Mental Health and Neuroscience, Maastricht, The Netherlands*
{*leonid.fedorov, martin.giese*}*@uni-tuebingen.de; joris.vangeneugden@gmail.com*

Keywords: Action Recognition, Multistable Perception, Biological Motion, Neural Fields, Shading

Abstract: Body motion perception from impoverished stimuli shows interesting dynamic properties, such as multistability and spontaneous perceptual switching. Psychophysical experiments show that such multistability disappears when the stimulus includes also shading cues along the body surface. Classical neural models for body motion perception have not addressed perceptual multistability. We present an extension of a classical neurodynamic model for biological and body motion perception that accounts for perceptual switching, and its dependence on shading cues on the body surface. We demonstrate that a set of psychophysical observations can be accounted for in a unifying manner by a hierarchical neural model for body motion processing that includes an additional shading pathway, which processes luminance gradients within the individual body segments. The goal of our model is to explain psychophysics and neural mechanism in the brain.

1 INTRODUCTION

The perception of body motion from image sequences requires the dynamic integration of complex spatio-temporal visual patterns. This important visual function is accomplished by processing within a hierarchy of cortical areas along the visual pathway. Psychophysical studies suggest depth cues are important for biological motion perception (Jackson and Blake, 2010). In absence of such depth information, e.g. in point-light walkers, body motion perception can become multistable (Vanrie and Verfaillie, 2004). Then the same stimulus can be perceived as alternating randomly between two interpretations that correspond to two different walking directions (Vanrie and Verfaillie, 2006). Multistabile phenomena has been also investigated in the context of static ambiguous figures and binocular rivalry (Leopold and Logothetis, 1999), (Blake and Logothesis, 2001), as well as in structure from motion (Andersen and Bradley, 1998). An example of the body motion stimulus that produces such multistability is shown in Fig. 1A (panel SILHOUETTE). For this stimulus, an articulating silhouette without intrinsic shading cues, observers perceive the walker alternately walking obliquely into or out of the image plane. The two reported percepts correspond to the unambiguous walking directions indicated in panels TOWARDS and AWAY. The figure illustrates also that this perceptual ambiguity disappears when shading gradients are added to the surface of the walker, which provide information about the surface orientation of the body segments and occlusions.

Existing physiologically-inspired neural models for the processing of body motion and goal-directed actions (e.g. (Giese and Poggio, 2003), (Lange and Lappe, 2006), (Escobar and Kornprobst, 2008), (Jhuang et al., 2007), (Fleischer et al., 2013) and (Layher et al., 2014)) do not reproduce such multistability, or at least never have investigated this phenomenon. Computer vision and deep learning architectures for body motion recognition do not address perceptual multistability. Thus, the study of such phenomena is important for neuroscience, even if such multistability is often unwanted in technical action recognition systems.

In the context of low-level vision, perceptual multi-stability and the underlying neural dynamics have been extensively studied e.g. in the context of binocular rivalry (see e.g. (Wilson, 2003)), visual motion integration (Rankin et al., 2014), or as general property of attractor neural networks (Pastukhov et al., 2013).

The goal of this paper is to extend existing physiologically-inspired neural models (not computer

vision algorithms) in a way that accounts for multi-stability in action perception, where we use as example an established model that has been shown to account jointly for many experimental results in this area (Giese and Poggio, 2003). We extend it in two ways: 1) by introduction of a multi-dimensional neural field that accounts for multi-stable behavior by lateral interactions between shape-selective neurons; 2) by addition of a new pathway that realizes robust processing of intrinsic luminance gradients along the surface of the body segments.

The paper is structured as follows: after discussing related work in the following section, we describe the developed architecture in section 3. In section 4 we show simulation results, illustrating that the model provides a unifying account for several key psychophysical results, followed by a brief discussion in section 5.

2 RELATED THEORETICAL WORK

Body motion recognition has been a core topic in computer vision and many technical neural architectures for this purpose have been proposed (Edwards et al., 2016), (Nguyen et al., 2016), (Ziaeefard and Bergevin, 2015), (Lee et al., 2014). The goal of that work is typically a maximization of recognition performance, not a reproduction of perceptual dynamics of humans. This paper does not contribute to computer vision or machine learning and is entirely focused on modeling of the brain.

We follow the approach in physiologically-plausible models of body motion perception, such as (Giese and Poggio, 2003), (Lange and Lappe, 2006), (Escobar and Kornprobst, 2008), (Fleischer et al., 2013), (Layher et al., 2014), while other biological models in this area (e.g. (Thurman and Lu, 2014) (Thurman and Lu, 2016)) account for experimental data without direct relationship to neural mechanisms.

Diverse approaches (see (Tyler, 2011)) have been proposed for the analysis of shape from shading, but typically not related to the processing of body motion. Perceptual dynamics and perceptual switching have been extensively studied in the context of low-level vision (reviews see e.g. (Leopold and Logothetis, 1999), (Sterzer et al., 2009), (Pastukhov et al., 2013)). Multistability in the processing of non-rigid motion has been rarely studied in neural modeling.

While hierarchical technical algorithms in computer vision typically focus on the problem how the body motion patterns (e.g. the direction of body movement) might be distinguished, our model tries to unify this account with a reproduction of the dynamics of perceptual organization in humans which emerges specifically for the SILHOUETTE stimulus, where for the same stimulus two alternating percepts emerge. This problem is typically not addressed in technical recognition systems, and to our knowledge no account for this phenomenon has been given in biologically-inspired neural models for motion recognition.

3 MODEL ARCHITECTURE

Our model builds on a previous neural model (Giese and Poggio, 2003), which has been shown to provide a unifying account for a variety of experimentally observed phenomena in body motion perception including physiological, psychophysical and fMRI data. The original model included a motion and a form pathway, processing shape and optic flow features. The pathways consist of a hierarchy of feature detectors that mimic properties of real cortical neurons. For the implementation in this paper we used only the form-pathway and extended it by a multi-dimensional neural field, and a new pathway for the processing of intrinsic luminance gradients. An extension by inclusion of an additional motion pathway is straight-forward, and will be part of future work.

3.1 Silhouette Pathway

The backbone of our model is a 'silhouette pathway' (Fig. 1B) that is identical to the the form pathway of the classical model (Giese and Poggio, 2003). Due to space limitations, we sketch here only some basics about this pathway and refer to the original publication (Giese and Poggio, 2003) with respect to details. In brief, the form pathway consists of a hierarchy of layers that process form features of increasing complexity along the hierarchy. More complex features are formed by combination of the features from previous layers. Levels that increase feature complexity are interleaved by layers that increase position and scale invariance by MAX pooling. The highest level of this shape processing hierarchy is formed by radial basis function units (called 'snapshot neurons') that have been trained with the feature vectors that correspond to keyframes from training movies showing the recognized action. Each snapshot neuron responds selectively to the body posture that corresponds to time instance θ (within the gait cycle). In addition, consistent with physiological data (Vangeneugden et al., 2011), we assume these neurons are view-specific, where the variable ϕ specifies the preferred view an-

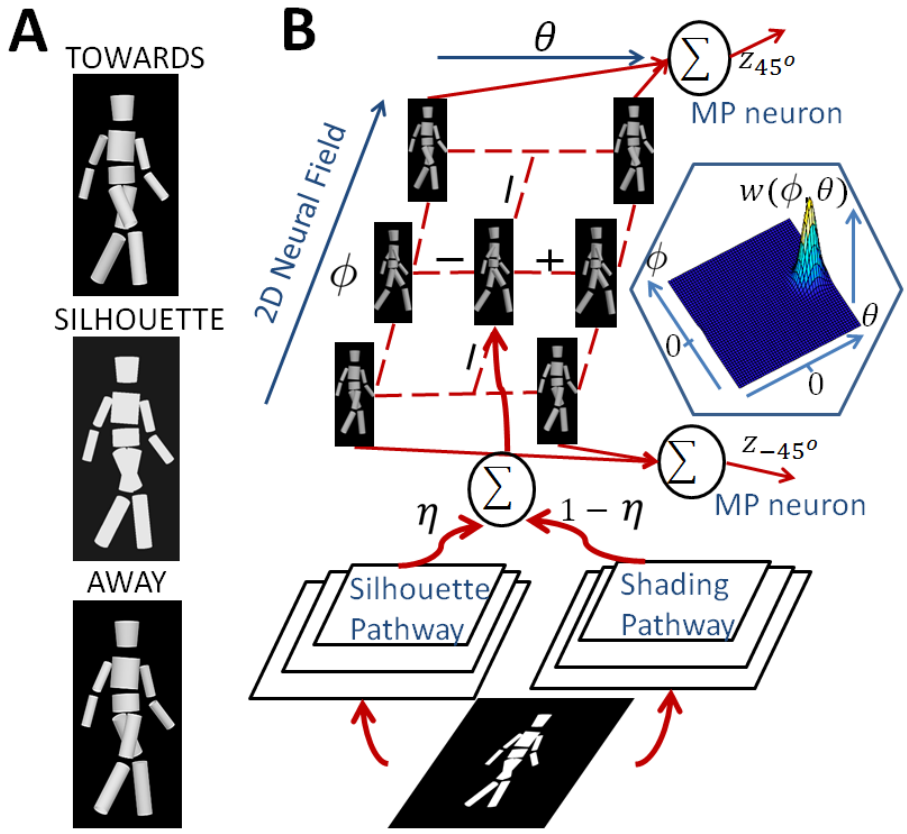

Figure 1: A. Snapshots from movies showing dynamic walker: TOWARDS shaded walker, walking direction 45 deg; SILHOUETTE bistable silhouette walker and AWAY shaded walker, walking direction -45 deg. B. Model architecture. Stimulus is analyzed by Silhouette and Shading pathways. Their outputs are linearly combined and mapped linearly onto the input of a 2D dynamic neural field that consists of laterally coupled snapshot neurons. Inset shows the lateral interaction kernel of the field. The field activity is read out by Motion Pattern (MP) neurons that encode the perceived walking directions ± 45 deg.

gle of the neuron. (We assume that the side view of a walker walking to the right in the image plane defines the view direction $\phi = 0$). Very similar architectures underlie many other classical and modern neural and deep models for object recognition, where the popular deep architectures are typically trained with much more data and often include many more layers. Since the goal of this paper is to model the perceptual dynamics, and not to maximize recognition rate, we used this simple hierachical model, where extension with modern deep architectures as front-end seem straight-forward.

3.2 Shading Pathway

The described simple form pathway recognizes body shape on backgrounds with sufficient contrast. However, it turned out that with small amounts of training data it is difficult to accomplish with this architecture a robust recognition of the silhouette shape together with a high sensitivity for the luminance shading gradients that disambiguate the depth structure. As one possible solution to this problem we implemented a second pathway that is specialized for the processing of intrinsic shading gradients using physiologically-plausible operations (Fig. 1B). We do not claim this is the only possible solution, but it is one that works with small amounts of training data.

The first level of this new pathway overlaps with the first hierarchy level of the silhouette pathway, described above. It consists of Gabor filters that are selective for local orientation features at different positions, and for different spatial scales. Let $G_{\mathrm{e,u}}(x, y, \alpha, \sigma)$ signify the output signal of the even (e) or uneven (u) Gabor filter with preferred position (x, y), preferred orientation α (we used 8 orientations), and scale σ (we used 1 scale for the given small stimuli set). The activations of the uneven Gabor filters provide a population code for the local luminance gradients.

By pooling of the responses of the Gabor filters with the same preferred position over all orientations we obtain position-specific detectors for contours with the output signals:

$$C(x,y) = \max_{\{\mathrm{e,u}\},\alpha,\sigma} |G_{\mathrm{e,u}}(x,y,\alpha,\sigma)|. \quad (1)$$

This output signal was used to suppress the responses of the uneven Gabor filters along the external contour of the body, exploiting multiplicative gating. The outer contour of the body typically creates strong local contrast that dominates the detector responses, so that the weak intrinsic gradients that signal the 3D structure cannot be reliably estimated from the neural responses. A population vector signaling the intrinsic luminescence gradients is given by the gated signal:

$$L(x,y,\alpha,\sigma) = [G_{\mathrm{u}}(x,y,\alpha,\sigma) \cdot H(\lambda_1 - C(x,y))]_+. \quad (2)$$

Here λ_1 is a positive constant, and the function $H(x)$ is the Heaviside function, thus $H(x) = 1$ for $x > 0$ and $H(x) = 0$ otherwise.

The next level of the shading pathway consists of (partially) position-invariant detectors for local luminance gradients. Their responses are computed by pooling of the gated responses of gradient detectors for the same preferred gradient direction α over all positions and scales in a quadratic neighborhood $\mathcal{U}(x',y')$ of the point (x',y') using a maximum operation, providing the output signals:

$$D(x,y,\alpha) = \max_{(x,y)\in\mathcal{U}(x',y'),\sigma} L(x,y,\alpha,\sigma). \quad (3)$$

These position-invariant detectors were defined for substantially less spatial positions, resulting in a strong spatial down-sampling (6,480,000 position- and scale-specific detectors vs. 648 position-invariant detector units).

In order to make recognition robust against fluctuating weak features, we selected the strongest features that provide input to the radial basis function units. We selected those features that showed the maximum variance over the training data (where clearly much more sophisticated feature selections are available that might lead to better results). We computed the circular variance of the detectors at position (x,y), exploiting the (complex) circular mean:

The (complex) circular mean of these responses is given by:

$$m(x,y) = (1/K) \sum_{k=1}^{K} \sum_{\alpha} D^{(k)}(x,y,\alpha) \exp(i\alpha), \quad (4)$$

where K is the number of training patterns. A circular variance measure is then given by the formula:

$$V(x,y) = \sum_{k=1}^{K} \left| \sum_{\alpha} D^{(k)}(x,y,\alpha) \exp(i\alpha) - m(x,y) \right|. \quad (5)$$

We selected the direction-specific responses $D(x,y,\alpha)$ that fulfilled the relationship:

$$V(x,y) > \lambda_2, \quad (6)$$

where $\lambda_2 > 0$ is a threshold parameter. In total 9 out of 81 feature vectors were selected according to this criterion.

The next level of the shading pathway is formed by Gaussian radial basis functions, whose centers were trained with the feature vectors $\mathbf{p}^l$ (including only the selected features) that were generated by individual keyframes from the training movies. For the results shown here, the shading pathway was trained with movies of fully shaded walkers, shown with view directions -45 deg and 45 deg. In other implementations, we have realized such models with a continuum of different views (Fleischer et al., 2013).

The RBF network returns an 50-dimensional output vector $\mathbf{R}_{\mathrm{SH}}(t)$ for each keyframe at time t, where the components of this vector are given by:

$$R^l_{\mathrm{SH}}(t) = \exp(-\lambda_3 ||\mathbf{p}(t) - \mathbf{p}^l||^2), \quad (7)$$

where $\mathbf{p}(t)$ is the feature vector for the actual input frame, and where the components correspond to the different keyframes and associated training views.

In order to link the shape recognition pathway to dynamic neurons that reproduce the perceptual dynamics, the outputs of the RBF units were mapped linearly onto a discretely sampled two-dimensional input activity distribution $s_{\mathrm{SH}}(\theta,\phi;t)$ that provides input to the neural field that is described below. Signifying by $\mathbf{s}_{\mathrm{SH}}(t)$ the appropriately reordered sampling points, the linear mapping was given by the equation:

$$\mathbf{s}_{\mathrm{SH}}(t) = \mathbf{W}(t)\mathbf{R}_{\mathrm{SH}}(t). \quad (8)$$

The weight matrices $\mathbf{W}(s)$ were learned by ridge regression from a training set that consisted of pairs of vectors $\mathbf{R}_{\mathrm{SH}}(t)$ for each training keyframe, and a corresponding vector $\mathbf{s}_{\mathrm{SH}}(t)$ that was computed from an idealized two-dimensional input activity distribution $s_{\mathrm{SH}}(\theta,\phi;t)$. The idealized activity distribution was given by a Gaussian peak that was centered at the keyframe number θ and the corresponding view ϕ of the walker (s.b.). A similar input distribution $s_{\mathrm{SL}}(\theta,\phi;t)$ was computed by a corresponding linear mapping in the silhouette pathway. The total input distribution of the neural field was then computed by 'cue fusion', modeled by a convex combination

of two input distribution functions according to the equation:

$$s(\theta,\phi;t) = \eta s_{\mathrm{SL}}(\theta,\phi;t) + (1-\eta)s_{\mathrm{SH}}(\theta,\phi;t), \quad (9)$$

with $0 \leq \eta \leq 1$. Choosing $\eta = 1$ one can eliminate the influence of the shading pathway.

3.3 Dynamic Neural Field of Snapshot Neurons

The core of our model is a dynamic recognition layer that is implemented as a two-dimensional neural field of Amari type (Amari, 1977), which consists of body shape-selective neurons that are laterally connected (Fig. 1B). Consistent with physiological data (Vangeneugden et al., 2011), we assume that such neurons encode body shapes that emerge during actions in a view-specific manner. In the spatial continuum limit, we can describe the activity of neurons encoding the body shape that corresponds to the normalized time θ ($0 \leq \theta \leq 2\pi$) during the gait cycle and the view angle ϕ by the function $u(\phi,\theta,t)$. The network dynamics is given by the equation ($\star$ signifying a spatial convolution):

$$\tau_u \frac{d}{dt}u(\phi,\theta,t) = -u(\phi,\theta,t) + w(\phi,\theta) \star H(u(\phi,\theta,t)) + s(\phi,\theta,t) - h + \xi(\phi,\theta,t) - c_a a(\phi,\theta,t). \quad (10)$$

The input signal s was described above. For the trained stimulus movies it corresponds to an activity maximum that moves in θ-direction along the field. The lateral connectivity is specified by the interaction kernel $w(\phi,\theta)$ (whose shape is indicated by the inset in Fig. 1B). It stabilizes a traveling pulse solution in θ-direction and realizes a winner-takes-all competition in the ϕ-direction. As consequence, if multiple views are consistent with the stimulus, one view is selected by competition. The positive parameters τ_u and h define the time scale and the resting potential of the field. The variable $\xi(\phi,\theta,t)$ defines a Gaussian noise process whose statistics was coarsely adapted to the noise correlations from cortical data (Giese, 2014). These fluctuations essentially drive the perceptual switching in the model. Since action perception shows adaptive properties, such as high-level after-effects and fMRI adaptation, we also included a neural adaptation process in the model, which reduces the activity of snapshot neurons after extended firing. The corresponding adaptation variable follows the dynamical equation:

$$\tau_a \frac{d}{dt}a(\phi,\theta,t) = -a(\phi,\theta,t) + H(u(\phi,\theta,t)). \quad (11)$$

The positive constant c_a determines the strength of adaptation(τ_α is the time constant). The parameters of this adaptation dynamics were fitted to experimental data (Giese, 2014).

The activity of the neurons in the neural field was read out by motion pattern (MP) neurons, which signal the walking directions perceived in this case as AWAY from and TOWARDS the observer. These neurons compute the maximum of the neural field activity function $u(\phi,\theta,t)$ over the domains $\phi > 0$ and $\phi < 0$ in the (ϕ,θ) space, producing the output signals z_{45} and z_{-45}.

4 SIMULATION RESULTS

Testing the model after training with a non-shaded walker as illustrated in Fig. 1A, 1B and 1C, the output of the shading pathway remained silent because of the absence of intrinsic luminance gradients in this stimulus. The silhouette pathway was activated in an ambiguous way by this stimulus because the stimulus is consistent with walking in the directions $\pm$45 deg relative to the image plane. Consistent with simulations described in (Giese, 2014), this stimulus leads to a bistable solution of the neural field that alternates between two traveling pulse solutions that encode the spontaneous perceptual switching of a traveling pulse between the view angles $\phi = \pm 45$ deg (perception of TOWARDS or AWAY from the observer). In this case, the probabilities of the two percepts are almost identical (Fig. 2B). More detailed simulations show that the model coarsely reproduces also the switching time statistics of human perception, comparing it with experimental data (not yet published (Vangeneugden et al., 2012)). Fig. 2G shows a histogram of the percept times for the model, and Fig. 2H the percept times estimated in the psychophysical experiment.

For shaded stimuli (see Figs. 1A TOWARDS and AWAY), when both pathways are included ($\eta = 0.5$), the model successfully disambiguates the walking direction: For the AWAY stimulus (direction -45 deg) the output neuron for AWAY remains always activated while the output neuron for TOWARDS remains silent. If an TOWARDS stimulus is shown (direction 45 deg) the situation is reverse and the TOWARDS output neuron is always active (Fig.2 C and D).

If however the shading pathway is deactivated ($\eta = 0$) again perceptual switching occurs, since the

output of the silhouette pathway is ambiguous, resulting in equal percept probabilities for either direction. The silhouette pathway is not sufficiently sensitive to disambiguate the stimulus robustly based on the available luminance gradients intrinsic to the body segments (Fig.2 E-F). This demonstrates the necessity of the shading pathway in the chosen architecture for the disambiguation of the percept.

The model makes several verifiable experimental predictions in relation to the time course of the adaptation process. An example is illustrated in Fig. 1I that shows a diagram of a typical adaptation experiment to demonstrate after-effects in action perception. First, an unambiguous adaptation stimulus (TOWARDS or AWAY) is presented to participants, where the duration of the adaptor (2, 6, 10, 14, 18 or 22 gait cycles) was varied over different blocks of the experiment. After this stimulus (and a fixed Interstimulus Interval of 2.8 s) an ambiguous test stimulus (SILHOUETTE) is presented for 3 gait cycles, asking for the perceived walking direction.

The predicted results for such an experiment (from 20 repeated simulations) are presented in Fig. 1J, which shows the probabilities of the percept for the ambiguous test stimulus (which was identical in all cases). With increasing the duration of the adaptor stimulus the probability that participants perceive the test stimulus as walking in the same direction as the adaptor decreases. A significant decrease of the percept probability (from 0.5 without adaptator presentation) is already perceived for the shortest adaptor duration of 2 gait cycles, and we observed a further decrease with longer adaptor durations (where 1 gait cycle corresponds to 1.4 seconds of stimulus duration).

This behavior is consistent with after-effects, as investigated previously for many modalities (motion, lightness, etc) in low-level vision. Such after-effects for action perception with a similar time course have been shown for other types of action stimuli in the literature (see (Barraclough and Jellema, 2011), (de la Rosa et al., 2014)), and we are presently running psychophysical experiments to verify this prediction of the model in detail.

A further set of experiments that we are presently running, and for which the model provides quantitative predictions, investigates the interdependence of the stability of action percepts and the switching times between the different percepts (which depend on the mean-first passage times of the corresponding attractors). This extends studies that have been made for muti-stability of low-level motion perception (Hock et al., 1993) to the domain of action perception.

5 CONCLUSIONS

To our knowledge, we have described the first biologically-inspired neural model that accounts simultaneously for the following properties of body motion perception: (i) perceptual multi-stability and switching, (ii) switching time statistics and (iii) the influence of shading information on the perceptual dynamics. We showed that the model reproduces the psychophysically observed phenomenology and distributions of the percept times. Since the model is based on learned templates, these results would transfer trivially to other action patterns with the similar form of bistability in the view domain.

It is important to stress that the goal of this paper was the modeling of the perceptual dynamics, and neither the proposal of novel deep shape or action recognition architecture, nor the claim that the proposed two-pathway architecture is significantly better for shape recognition. Testing this claim would require additional experiments with larger data sets, and was not the focus of this paper. Also it remains to be shown whether any of the popular recurrent deep architectures reproduce the details of the human perceptual dynamics.

Future work will have to extend the model for more stimuli and include more accurate fits of experimental data.

ACKNOWLEDGEMENTS

The first author thanks Tjeerd Dijkstra for his insightful commentary on the analysis of the Amari field behavior. Funded by: BMBF, FKZ: 01GQ1002A, ABC PITN-GA-011-290011, CogIMon H2020 ICT-644727; HBP FP7-604102; Koroibot FP7-611909, DFG GZ: KA 1258/15-1; HFSP RGP0036/2016.

REFERENCES

Amari, S. (1977). Dynamics of pattern formation in lateral inhibition type neural fields. *Biological Cybernetics*.

Andersen, R. and Bradley, D. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Sciences*.

Barraclough, N. and Jellema, T. (2011). Visual aftereffects for walking actions reveal underlying neural mechanisms for action recognition. *Psychological Science.*

Blake, R. and Logothesis, N. (2001). Visual competition. *Nature Review Neuroscience*.

de la Rosa, S., Streuber, S., Giese, M., Buelthoff, H., and Curio, C. (2014). Putting actions in context: Visual

action adaptation aftereffects are modulated by social contexts. *PLOS One*.

Edwards, M., Deng, J., and Xie, X. (2016). From pose to activity. In *Computer Vision and Image Understanding*. Elsevier Science Inc.

Escobar, M. and Kornprobst, P. (2008). Action recognition with a bioinspired feedforward motion processing model: the richness of center-surround interactions. In *ECCV'08. 10th European Conference on Computer Vision*. Springer Berlin Heidelberg.

Fleischer, F., Caggiano, V., Thier, P., and Giese, M. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience*.

Giese, M. (2014). Skeleton model for the neurodynamics of visual action representations. In *Artificial Neural Networks and Machine Learning ICANN 2014, Lecture Notes in Computer Science*. Springer International Publishing.

Giese, M. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements and action. *Nature Reviews Neuroscience*.

Hock, H., Kelso, J., and Schoener, G. (1993). Bistability and hysteresis in the perceptual organization of apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*.

Jackson, S. and Blake, R. (2010). Neural integration of information specifying human structure from form, motion, and depth. *Journal of Neuroscience*.

Jhuang, H., Serre, T., Wolf, L., and Poggio, T. (2007). A biologically inspired system for action recognition. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE.

Lange, J. and Lappe, M. (2006). A model for biological motion perception from configural form cues. *Journal of Neuroscience*.

Layher, G., Giese, M., and Neumann, H. (2014). Learning representations of animated motion sequencesa neural model. In *Topics in Cognitive Science*. Topics in Cognitive Science.

Lee, T., Belkhatir, M., and Sanei, S. (2014). A comprehensive review of past and present vision-based techniques for gait recognition. In *Multimedia Tools and Applications*. Kluwer Academic Publishers.

Leopold, D. and Logothetis, N. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Science*.

Nguyen, D., Li, W., and Ogunbona, P. (2016). Human detection from images and videos. In *Pattern Recognition*. Elsevier Science Inc.

Pastukhov, A., Garca-Rodrguez, P., Haenicke, J., Guillamon, A., Deco, G., and Braun, J. (2013). Multi-stable perception balances stability and sensitivity. *Frontiers in Computational Neuroscience*.

Rankin, J., Meso, A., Masson, G. S., Faugeras, O., and Kornprobst, P. (2014). Bifurcation study of a neural field competition model with an application to perceptual switching in motion integration. *Journal of Computational Neuroscience*.

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Science*.

Thurman, S. and Lu, H. (2014). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience*.

Thurman, S. and Lu, H. (2016). A comparison of form processing involved in the perception of biological and nonbiological movements. *Journal of Vision*.

Tyler, C. (2011). *Computer Vision: From Surfaces to 3D Objects*. Chapman & Hall/CRC, London, 1st edition.

Vangeneugden, J., de Maziere, P., van Hulle, M., Jaeggli, T., van Gool, L., and Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *Journal of Neuroscience*.

Vangeneugden, J., van Ee, R., Verfaillie, K., Wagemans, J., and de Beeck, H. (2012). Activity in areas mt+ and eba, but not psts, allow prediction of perceptual states during ambiguous biological motion. In *Society for Neuroscience Meeting*. Society for Neuroscience.

Vanrie, J. and Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, and Computers*.

Vanrie, J. and Verfaillie, K. (2006). Perceiving depth in point-light actions. *Perception and Psychophysics*.

Wilson, H. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences*.

Ziaeefard, M. and Bergevin, R. (2015). Semantic human activity recognition: A literature review. In *Pattern Recognition*. Elsevier Science Inc.

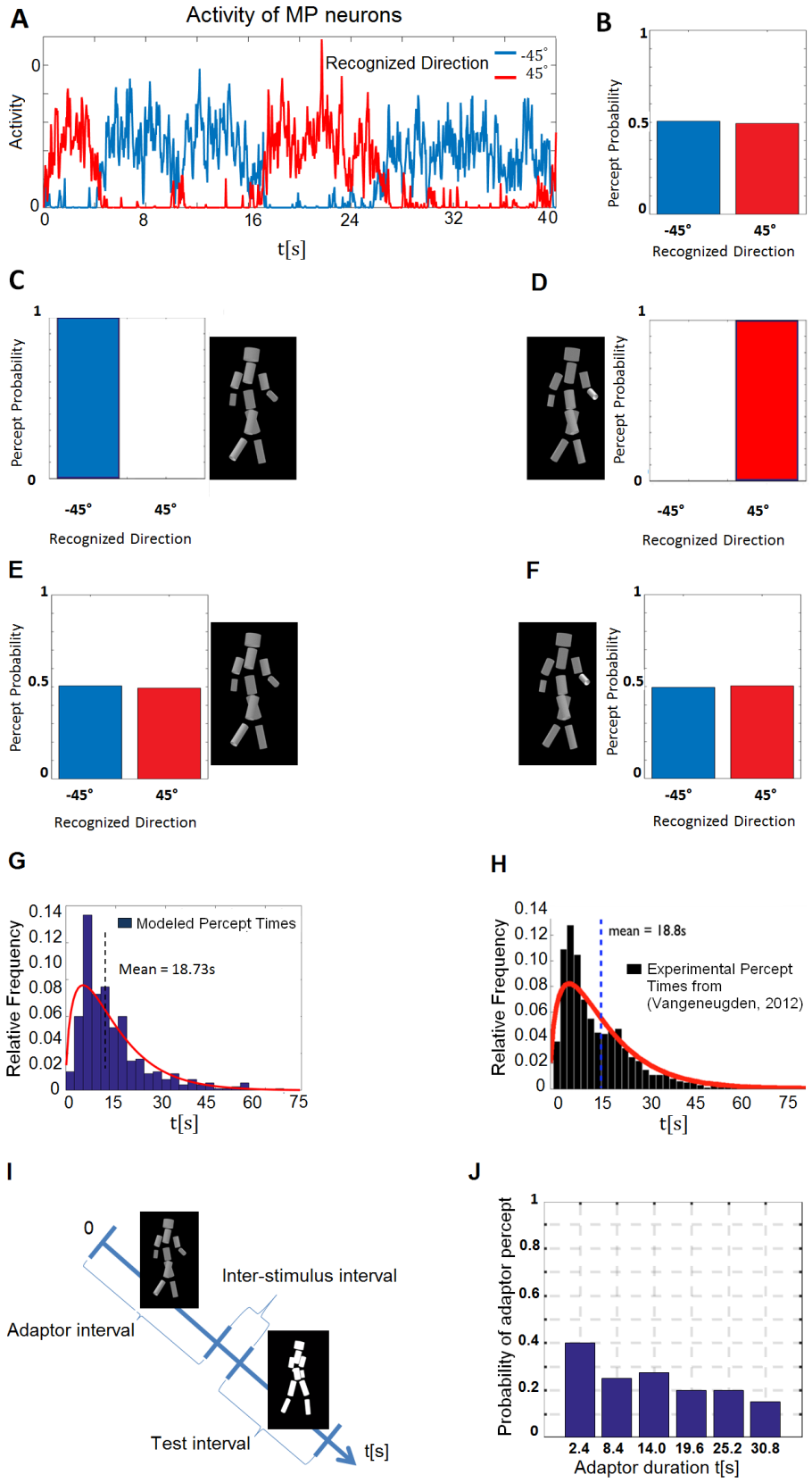

Figure 2: A. Time courses of the activity of motion pattern neurons for depth-ambiguous walker stimulus. B-F. Percept probability of the motion pattern neurons for the percepts TOWARDS and AWAY for (B) depth-ambiguous walker for model with both pathways ; (C) shaded −45 deg (AWAY) walker for model with both pathways; (D) same for shaded 45 deg (TOWARDS) walker; (E) shaded 45 deg (TOWARDS) walker for model without shading pathway; (F) same for shaded 45 deg (TOWARDS) walker; G-H. Histogram of percept times (PT) from experimental data (Vangeneugden et al., 2012) and from the model. I. Paradigm for testing after-effects in action perception which is compatible with our model. After presentation of an unambiguous adaptor stimulus (AWAY or TOWARDS), and a fixed Inter-stimulus Interval, an ambiguous test stumulus (SILHOUETTE) is presented. J. Probability that test stimulus is perceived as walking in the adaptor direction as a function of the duration of the adaptor.