A causal inference model for the perception of complex motion in the presence of self-motion

Sabyasachi Shivkumar***, Gregory C. DeAngelis, Ralf M. Haefner***
Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA
$\star$ {sshivkum,rhaefne2}@ur.rochester.edu

Abstract

Our subjective percept of motion differs from the actual velocity on the retina in striking ways that have tradition-ally been studied by disparate fields. Here, we present a Bayesian model, as well as new data to support it, that unifies our understanding of motion perception of complex objects ("grouping") with our understanding of the influence of self-motion on our perception ("flow-parsing").

The central (recurring) motif in our hierarchical model is a prior over velocity consisting of a mixture of both a delta and a Gaussian centered on zero. This simple mod-ification of the classic slow speed prior implies a "causal inference" process over whether the object is stationary or moving. Applied to multiple visual elements it leads to a "chunking" of these elements into groups, and groups of groups, with the goal to make the relative speed of as many of the elements zero with respect to the group they are inferred to belong to. As a result, our model infers in-dividual motion relative to a group, and accounts for any inferred self-motion based on optic flow.

Preliminary data from two experiments confirm new predictions of the model.

Keywords: motion perception; causal inference; flow parsing; hierarchical motion

Introduction

A long line of empirical work has established that our per-ception of motion is both hierarchical (Johansson, 1950): the motion of visual elements that the brain infers to belong to a group of other elements (e.g. an object) are perceived to be relative to the motion of that object. Such relative perception can even be nested, e.g. the motion of a part of an object is perceived to be relative to the motion of an object that itself is perceived to move relatively to an even larger group of objects. Gershman and colleagues recently proposed a hierarchical Bayesian model that can correctly infer both groupings of vi-sual elements and the perceived relative motion (Gershman, Tenenbaum, & Jäkel, 2016). In our work we modify and extend Gershman's model in two key ways. First, we reformulate it in a way that replaces its ad-hoc simplicity prior based on a Chi-nese restaurant process by a slow-speed mixture prior that is justified by the statistics of natural inputs. Second, we extend it to account for self-motion and vestibular inputs and thereby enable it to account for a large body of data describing hu-man motion perception in the presence of optic flow (Warren & Rushton, 2009).

Central Model Motif for Motion Perception

Model Description: First consider the case of an object that is part of a larger object ("group") moving in the world. In or-der to determine the velocity of this object ($\vec{v}^{\text{object}}$), the subject has two sources of information: (a) The observed velocity of the object as manifested on its retina ($\vec{o}^{\text{object}}$). And (b) the pre-dicted velocity of the object based on knowledge of the larger object that it belongs to ("group") having velocity ($\vec{v}^{\text{group}}$). We hypothesize that the subject combines these two sources of information by performing inference in the generative model in Figure 1A. In this generative model, the object velocity ($\vec{v}^{\text{object}}$) is modeled as a sum of the group the object belongs to ($\vec{v}^{\text{group}}$) and the object's velocity relative to the group ($\vec{v}^{\text{object}}_{\text{group}}$) with some uncertainty associated with approximate computations ($\Sigma^{\text{estimation}}$). Eg: the velocity of a person moving on a train is the velocity of the train and the person's velocity relative to the train. The observed object velocity ($\vec{o}^{\text{object}}$) is modeled as $\vec{v}^{\text{object}}$ corrupted with visual sensory noise ($\Sigma^{\text{estimation}}$). The prior velocities for $\vec{v}^{\text{group}}$ and $\vec{v}^{\text{object}}_{\text{group}}$ are modeled as mixture of a delta at 0 and a normal distribution with mean 0 and vari-ance Σ^{prior} illustrated in Figure 1B. We hypothesize that the subject perceives the object velocity as the inferred $\vec{v}^{\text{object}}_{\text{group}}$ if it is non-zero, i.e. the object moves relative to the group else perceives it as the inferred group velocity ($\vec{v}^{\text{group}}$).

Link to slow speed prior: The mixture prior used in our model is different from the traditional slow-speed prior (Gaus-sian centered on zero) used in earlier studies (Stocker & Si-moncelli, 2006). We believe that this form of mixture better describes the statistics of the outside world, and hence the brain's expectations over velocities in the world since a sub-stantial fraction of objects in the world are actually stationary, while a Gaussian prior without the delta at zero would predict that almost all objects move with some non-zero velocity.

Causal Inference: The delta component of the prior allows the model to group different motion elements into a hierarchy of motion groups. The reason is that multiple points moving with similar velocity have a higher likelihood under the model of being part of an object which moves with that velocity in-stead of being separate elements with non-zero velocity each. It is straightforward to show that this process of grouping fol-lows a causal inference mechanism analog to that previously proposed for multi-sensory inference (Körding et al., 2007). In our model, the mixture prior allows the subject to infer which group an object belongs to and to infer the velocity of the ob-ject relative to that group.

Complex Motion Perception: This motif can be recur-sively extended to model the motion perception at different hi-

erarchical levels from simple "dot" elements to groups of dots, groups of dots to supergroups of groups and so on (Figure 1C). Furthermore, it is straightforward to extend this model to incorporate self-motion (Figure 1C left) the prior over which we again model as a mixture of a delta at zero and a Gaussian. This choice again implies a causal inference process over whether the subject is stationary or not, mathematically equivalent to that recently proposed by (Dokka, Park, Jansen, DeAngelis, & Angelaki, 2019). In the next sections, we adapt this general model to predict the motion perception of complex objects and motion perception under perceived self motion for two specific experiments designed to test its novel elements: the mixture prior implying causal inference, and the combination with self-motion.
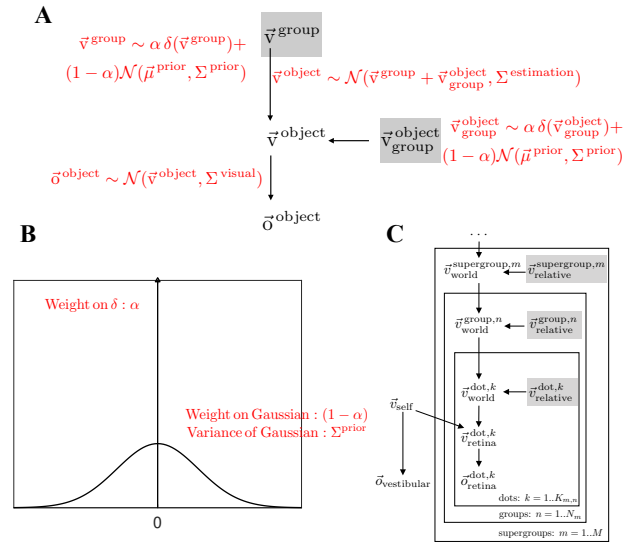

Figure 1: Central motif of our generative model for motion perception: (A) Generative model for object velocity as the sum of the group velocity the object belongs to and the velocity of the object relative to the group. Perceived object velocity is modeled as the lowest inferred non-zero shaded variable. (B) Mixture prior consisting of a delta and a Gaussian. Inference using such a prior leads to grouping and causal inference which have been observed in motion perception. (C) Complete hierarchical generative model including self-motion.

Hierarchical Motion Perception

In this section, we study how our perception of object velocity is affected by the presence of other objects leading to it being grouped together with them. We adapt the generative model described in the previous section to predict behavior.

Task Description: The schematic for the task is shown in Figure 2A (top). The subject fixates at the center, observes the object (red dot) move at a particular angle, θ with respect to the horizontal (Figure 2A, bottom) and reports the perceived direction of movement with a dial. The red dot is surrounded

by $N_{\text{dots}}^{\text{group}}$ green dots which move horizontally concurrently with the red dot ($N_{\text{dots}}^{\text{group}}$ is chosen as 1,2,3,5 or 10). All dots move back and forth once before moving in the direction which the subject has to report (indicated by change in colour of the fixation dot). The trials with different $N_{\text{dots}}^{\text{group}}$ are interleaved with conditions where the grouping dots are stationary. The speed of the group dots are chosen such that they match the x component of the red dot velocity (Figure 2A, bottom).

Model Description: We model the subject's perception as inference in the generative model shown in Figure 2B with the coloured boxes indicating the possible candidates for perceived variables. The subject observe the velocity of all the dots ($\vec{o}$) which are modeled as the true dot velocity ($\vec{\varepsilon}$) corrupted by external (eg: screen resolution) and internal noise (eg: photoreceptor noise). All velocities are represented by their x and y components. The inferred velocity of each dot is modeled using the motif in terms of the velocity of a group and its velocity relative to the group. While the model captures the grouping due to common velocity components, objects can be part of different groups despite having a common velocity (e.g. a bird flying in front of a moving train). Thus, the subject also infers if different objects belong to a group before they can split the velocity as per the motif. Here, we model this causal inference by inference over $\vec{C}_{\text{group}}^{\text{object}}$ which indicates if the object is part of the group or not.

Predictions: We predict the distribution of subject responses over different trials from the model. When the number of moving green objects is 0 (stationary green dots), we predict that the subject perceives the veridical value (Top left panel). With concurrently moving green dots, we predict three modes for the subject percepts: (a) Veridical value of object velocity (identity line) if the subject does not infer the object and green dots to be part of the same group (b) group velocity (0 degrees) if the subject infers that the object moves as part of the group (c) relative velocity to the group (90 degrees since horizontal velocities are matched). The mode at zero (signature of causal inference) indicate a pull towards the group velocity which is not predicted by traditional vector subtraction models.

Results: Model predicted distributions are indicated using the red violin plot and the responses of an example subject is indicated by the blue dots in Figure 2C. The predicted distributions by the model largely agree with our earlier predictions. The model predicts more reports of relative velocity direction as the angle between the patch and the group increases. For small angles, the model predicts the mode at zero which strengthens with number of dots (higher chance of grouping with more group dots). The responses of the example subject qualitatively match the model predictions.

Object Motion Perception under Self Motion

In this section, we study at how our perception of object velocity is affected due to perceived self motion and adapt the motif described in the previous section to model subject behavior.
Task Description: The schematic for the task is shown in Fig-

ure 3A (top). The subject fixates at the centre and observes the object (green dot) move at a particular angle, θ to the horizontal on the retina (Figure 2A, bottom) and reports the direction of perceived direction of the object with a dial. The background consist of red dots which whose velocity simulates an optic flow indicate of self motion either towards or away from the screen. The number of dots is interleaved across trials in a session and is chosen as 1,10,100 or 1000. The self motion velocity is chosen in such a way that the optic flow vector at the object location matches the horizontal component of the object's retinal velocity. (Figure 3B)

Model Description: We model the subject's motion perception in the task as inference in the generative model in Figure 3C. The notation is similar to the model in the previous section with $\vec{o}$ indicating the observed velocity, $\vec{\varepsilon}$ indicating the true velocity and $\vec{v}$ indicating the inferred velocity. The subject infers their self motion velocity from their vestibular observations and the optic flow motion (visual) modeled using causal inference. The velocity of all dots on the retina is inferred as a sum of the dot's velocity in the world and the velocity due to the subject's self motion. This is calculated as the velocity of dot on the retina if it has been stationary in the world (denoted by $\vec{v}^{\text{stationary}}$ which depends on the subject's self motion and the position of the object in the world.

Predictions: The model predicts that subject's infers their self motion from the observed dot velocities. Since there is a mixture prior (Figure 1B) over the perceived self motion, the model predicts that the subject's percept of self motion increases with the number of dots. This happens as more dots moving with a velocity consistent with self motion have a higher likelihood of being 'actually' stationary in the world and a moving subject. The inferred self motion determines the predicted velocity of a stationary object in the world which is subtracted from the observed retinal velocity to determine the object's velocity in the world (Perceiving which can be modeled as "flow parsing"). The magnitude of flow parsing can be computed using the flow parsing gain which is ≤ 1 where 1 indicates complete flow parsing (percept of 90 degrees in the task). Hence we can use the flow parsing gain as an indicator of the increase in the inferred self motion by the subject.

Results: The subject reports of the perceived angle of the moving object is given in Figure 3E. The subject reports are biased due to the optic flow dots consistent with the model predictions. Since the model predicts increase of the flow parsing gain (linked to self motion percept) with the number of dots, we compute the gain for the example subject (Figure 3D). The flow parsing gain increases with the number of dots for both forward and backward motion as predicted by the model.

References

Dokka, K., Park, H., Jansen, M., DeAngelis, G. C., & Angelaki, D. E. (2019). Causal inference accounts for heading perception in the presence of object motion. *Proceedings of the National Academy of Sciences*, *116*(18), 9060–9065.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision research*, *126*, 232–241.

Johansson, G. (1950). *Configurations in event perception: an experimental study*. Unpublished doctoral dissertation, Almqvist & Wiksell.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(9), e943.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, *9*(4), 578.

Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology*, *19*(18), 1555–1560.
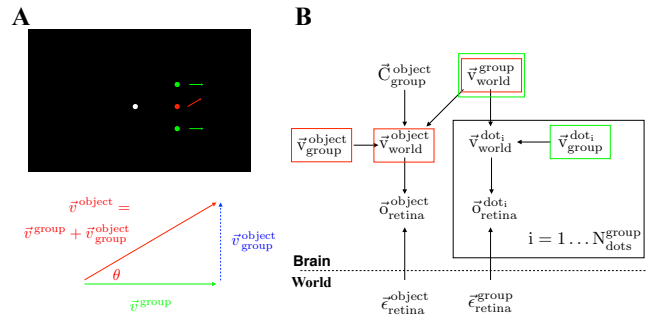
Figure 2: Hierarchical Motion Perception: (A) Task schematic. Subject has to report the perceived velocity direction of the red dot. The red dot is surrounded with varying number of green dots which move horizontally concurrently (B) Generative model hypothesized for modeling subject's perception. The model uses the motif described in Figure 1 to group different dots based on their velocity similarity (C) Model predictions for the distribution of subject responses in the task (red violin plot) and the responses of an example subject (Blue dots) on the task
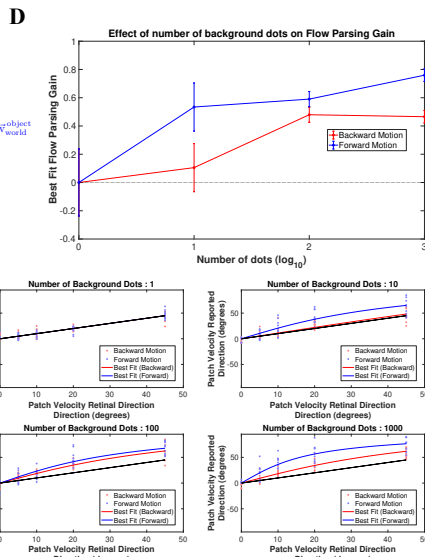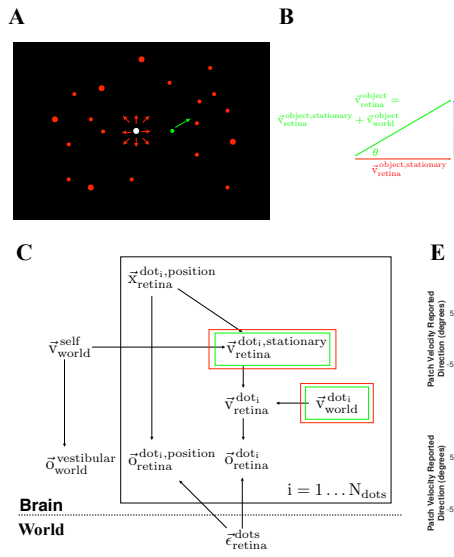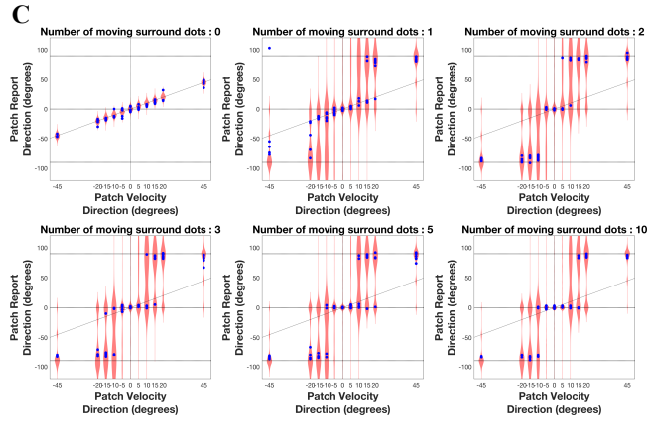

Figure 3: Object Motion Perception under Self Motion: (A) Task Schematic. Subject reports the perceived velocity direction of the green dot. Varying number of red dots constitute an optic flow inducing self motion either towards or away from the screen. (B) Velocities are chosen such that the optic flow vector at the object location matches the horizontal component of object velocity. Complete flow parsing in this experiment predicts a vertical percept. (C) Generative model hypothesized for modeling subject perception. The subject's inferred self motion affects the percept of the dot velocities. (D) Flow Parsing Gain: defined as the magnitude of self motion vector used in the inference of the object's velocity in the world. The gain increases with number of dots as predicted by the model. (E) Subject responses for different retinal velocity directions. The responses are biased by the optic flow dots (Flow parsing)